

ДЕРЖАВНИЙ УНІВЕРСИТЕТ ТЕЛЕКОМУНІКАЦІЙ
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ІНФОРМАЦІЙНИХ
ТЕХНОЛОГІЙ

Кафедра інженерії програмного забезпечення

Пояснювальна записка

до магістерської роботи

на ступінь вищої освіти магістр

на тему: «ПІДВИЩЕННЯ РЕЗУЛЬТАТИВНОСТІ РЕКЛАМНОЇ
КОМПАНІЇ НА ОСНОВІ МЕТОДІВ МАШИННОГО НАВЧАННЯ»

Виконав: студент 6 курсу, групи ПДМ-61

спеціальності 121 Інженерія програмного

забезпечення

(шифр і назва спеціальності)

Яновський Д.А.

(прізвище та ініціали)

Керівник _____

Жебка В.В.

(прізвище та ініціали)

Рецензент _____

(прізвище та ініціали)

Київ – 2022

ДЕРЖАВНИЙ УНІВЕРСИТЕТ ТЕЛЕКОМУНІКАЦІЙ

**НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ІНФОРМАЦІЙНИХ
ТЕХНОЛОГІЙ**

Кафедра Інженерії програмного забезпечення

Ступінь вищої освіти «Магістр»

Спеціалізація

ЗАТВЕРДЖУЮ

Завідувач кафедри
Інженерії програмного
забезпечення

О.В. Негоденко

“ _____ ” _____ 2022 року

З А В Д А Н Н Я
НА МАГІСТЕРСЬКУ РОБОТУ СТУДЕНТУ

Яновському Дмитру Андрійовичу

(прізвище, ім'я, по батькові)

1. Тема роботи: «Підвищення результативності рекламної компанії на основі методів машинного навчання».
Керівник роботи доцент кафедри, завідувач кафедри Технологій цифрового зв'язку Жебка В.В.,
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)
затвердженні наказом вищого навчального закладу від 12.10.2022 №122.
2. Строк подання студентом роботи 30.12.2022.
3. Вихідні дані до роботи:
 - 3.1. Науково-технічна література та дані рекламної компанії.
 - 3.2. Методи машинного навчання у галузі digital реклами.
4. Зміст розрахунково-пояснювальної записки (перелік запитань, які потрібно розробити):
 - 4.1. Порівняльний аналіз методів машинного навчання у аналізі даних.
 - 4.2. Методика для покращення результативності рекламної компанії на основі методу дерево рішень та її застосування.
 - 4.3. Дослідження шляхів забезпечення зростання ефективності рекламних компаній та збільшення рівня лояльності користувачів.

5. Перелік графічного матеріалу:
 - 5.1. Методи машинного навчання: дерево рішень.
 - 5.2. Методика підвищення ефективності рекламної компанії.
 - 5.3. Переваги та недоліки використання нейронних мереж в рекламі.
 - 5.4. Результат підвищення ефективності рекламної компанії
6. Дата видачі завдання 14.10.2022.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів магістерської роботи	Строк виконання етапів роботи	Примітка
1	Підбір науково-технічної літератури.	14.10.2022	Виконано
2	Дослідження методів машинного навчання.	21.09.2022	Виконано
3	Розробка методики для підвищення ефективності рекламної компанії за допомогою методів машинного навчання.	28.09.2022	Виконано
4	Розробка програмного забезпечення.	02.10.2022	Виконано
5	Дослідження та налаштування програмного забезпечення.	17.10.2022	Виконано
6	Оформлення роботи, розробка демонстраційних матеріалів	27.12.2022	Виконано
7	Здача роботи	31.12.2022	Виконано

Студент  Яновський Д.А.
 Керівник роботи _____ Жебка В.В.

РЕФЕРАТ

Текстова частина магістерської роботи 80 сторінок, 27 рисунків, 60 джерел.

МЕТОДИ МАШИННОГО НАВЧАННЯ; ДЕРЕВО РІШЕНЬ; ТАРГЕТИНГ;
РЕКЛАМНА КОМПАНІЯ; ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ.

Об'єкт дослідження – процес кластеризації цільової аудиторії для таргету.

Предмет дослідження – методи машинного навчання.

Мета роботи – оптимізація пошуку цільової аудиторії для рекламної компанії на основі методів машинного навчання.

Методи дослідження – методи машинного навчання, імітаційне моделювання, нейронна мережа.

Дане дослідження присвячене проблемі збільшення ефективності рекламних компаній на основі методів кластерного аналізу. З метою оптимізація пошуку цільової аудиторії для рекламної компанії у роботі проаналізовані існуюча класифікація та методи машинного навчання, які можна використати для збільшення рівня лояльності споживачів реклами. Здійснено обґрунтування створення методики класифікації на основі методу дерево рішень. Представлено впровадження застосування даної методики для збільшення ефективності рекламної компанії. Розроблено програмне забезпечення для аналізу даних, для подальшого підвищення лояльності цільової аудиторії.

Результати дослідження були використані для розробки програмного забезпечення для підвищення результативності рекламної компанії та впроваджені у робочий процес рекламного агентства.

ЗМІСТ

ВСТУП	8
1 ОСОБЛИВОСТІ ТА МЕХАНІЗМИ РЕКЛАМНОЇ КОМПАНІЇ	11
1.1 Особливості рекламної компанії в Україні та світі	11
1.2 Таргетинг – як механізм виділення цільової аудиторії	26
1.3 Постановка завдання дослідження	34
2 ВИБІР МЕТОДУ МАШИННОГО НАВЧАННЯ ДЛЯ ЕФЕКТИВНОЇ РЕКЛАМИ	36
2.1 Аналіз методів машинного навчання	36
2.2 Вибір та класифікація даних для машинного навчання	62
2.3 Розробка методики вибору цільової аудиторії	69
3 ПІДВИЩЕННЯ РЕЗУЛЬТАТИВНОСТІ РЕКЛАМНОЇ КОМПАНІЇ	86
3.1 Розробка методики підвищення результативності рекламної компанії	86
3.2 Аналіз ефективності запропонованої методики	88
3.3 Рекомендації для подальшого удосконалення рекламної компанії	90
ВИСНОВКИ	91
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	92
ДОДАТОК А	98

ВСТУП

Сьогодні світ стрімко розвивається. Замість бігбордів ми отримуємо гарні, великі екрани з динамічною інформацією. Замість газет тепер кожен має власний смартфон та може знайти будь-яку рід для своїх потреб. Так і в рекламі, вона також розвивається. З приходом соціальних мереж, життя кожного суттєво змінилося. Там ми проводимо свій вільний час, спілкуємося з друзями, переглядаємо фільми та навіть працюємо. Це привело до того, що були створені групи та бесіди, де адміністратори, чи всі користувачі можуть поділитися новиною та прокоментувати її.

Кожному приходили рекламні сповіщення та нам не зрозуміло звідки в рекламодавців контакти для зв'язку. Все просто: наприклад, продавець хоче знайти клієнта на ремонт авто. Достатньо буде знайти декілька груп, чи бесід з автомобільною тематикою у потрібному районі, зібрати всю аудиторію та розіслати їм рекламне оголошення. Але цей та схожі процеси мають безліч недоліків. Дивлячись на цю проблему, гіперкомпанії створили штучний інтелект, який зберігає дії користувача в Інтернеті, аналізує його поведінку та дозволяє рекламодавцям зменшити відсоток незацікавлених користувачів, оскільки оголошення відображаються тим людям, хто був пов'язаний з даною тематикою. Також є можливість фільтрувати аудиторію за віком, місцезнаходженням, типом та моделлю пристрою, а також за безліччю інших параметрів. Цей напрям називається таргетинг.

Але існування даної технології не достатньо для успішної рекламної компанії. Для гарного результату потрібні гарні фахівці, кожен з яких впевнено та якісно виконує поставлені задачі, а також потрібно аналізувати всі дані, проводити покращення медіа частини та зазвичай це забирає дуже багато сил та часу. Тому було вирішено створити методику для підвищення ефективності рекламної компанії за допомогою методів машинного навчання, оскільки велику частину роботи для обробки даних можна покласти саме на систему, котра здатна

аналізувати та самонавчатися й самовдосконалюватися. Такий підхід дозволить збільшити лояльність користувачів Інтернет до певної реклами, зменшити можливі витрати та збільшити дохід рекламної компанії, а також підвищити рівень її репутації. Цим і доводиться актуальність даного дослідження.

Мета роботи: оптимізація пошуку цільової аудиторії для рекламної компанії на основі методів машинного навчання.

Для досягнення цієї мети в роботі необхідно вирішити такі **завдання**:

1. Проаналізувати наукові праці з досліджуваної проблеми і обґрунтувати застосування машинного навчання для виділення цільової аудиторії.
2. Проаналізувати методи машинного навчання.
3. Обґрунтувати застосування методу дерево рішень.
4. Розробити та теоретично обґрунтувати методику для підвищення рекламної компанії на основі методу дерево рішень.
5. Розробити програмне забезпечення для кластеризації цільової аудиторії.

Виходячи з цього, об'єктом дослідження процес кластеризації цільової аудиторії для таргету, а предметом дослідження – методи машинного навчання.

Методи дослідження: методи машинного навчання, імітаційне моделювання, нейронна мережа.

Наукова новизна одержаних результатів: розроблено методику для підвищення ефективності рекламної компанії на основі методу дерево рішень.

Практичний результат: основні положення та результати магістерської роботи були використані для розробки програмного забезпечення для збільшення ефективності рекламної компанії та впроваджені у робочий процес рекламного агентства.

Апробація результатів дослідження:

1. Яновський Д.А. Методика підвищення результативності рекламної компанії на основі методів машинного навчання// ТІТ. №3, 2022, с. 25-30.
2. Яновський Д.А. Розробка методики вибору цільової аудиторії для підвищення результативності рекламної компанії // XV Науково-технічна

конференція «Сучасні інфокомунікаційні технології» . – Київ: ДУТ, 2022. – с. 93-95.

1 ОСОБЛИВОСТІ ТА МЕХАНІЗМИ РЕКЛАМНОЇ КОМПАНІЇ

1.1 Особливості рекламної компанії в Україні та світі

Кожна країна має свою культуру, традиції та особливості, тому компаніям, які працюють не тільки на локальному, а й на міжнародному ринку, необхідно підлаштовуватися під кожен менталітет окремо, щоб не виникало неприємних ситуацій і проблем. Зазвичай компанії розробляють типову рекламу для іноземних ринків, але така реклама не завжди ефективна, тому необхідно замовляти розробку реклами у фахівців, щоб вони могли глибоко вивчити певну країну і створити дійсно ефективний рекламний сервіс або продукт. Багато сучасних агентств вважають, що реклама, створена в різних країнах, має схожі характеристики та принципи виховання рекламних експертів. Насправді це не зовсім так, і аналіз багатьох компаній показує, що незалежно від того, де робиться реклама, у всіх країнах є свої способи та засоби донесення інформації до споживачів.

Більшість цих непорозумінь походить від міжнародної реклами. Нові ідеї людей, які займаються створенням реклами, зазвичай з'являються в повсякденному житті, і їм сприяє їхнє рідне місто чи оточення. Зібрати кілька людей в одному місті Європи для створення ефективного загальноєвропейського рекламного агентства навряд чи вдасться, оскільки таке агентство більше підійде для їхньої країни. Однак якщо ви збираєте кілька команд із різних європейських країн і замовляєте розробку реклами, яка їм підходить, у цьому випадку продукт виходить на ринок, і це може бути геніальною ідеєю. Абсолютно в усіх країнах є певні обмеження, тому іноді великим компаніям, які вирішують розробляти рекламу, потрібно зустрітися та враховувати всі характеристики людей.

Враховуючи, що розвиток реклами в різних країнах відбувається різними темпами через інновації, технології, концепції та принципи, лідером завжди буде

високорозвинена країна, незалежно від того, про яку країну йде мова. Глобальна реклама в США та країнах Європи завжди на високому рівні.

Реклама часто відображає суспільство і завжди слідує останнім тенденціям. Маркетологи використовують його як інструмент для специфічних емоційних реакцій. Щоб вплинути на аудиторію, бренди поєднують сильні заголовки, добре досліджену ідею та ефективні методи доставки. Synovate, міжнародна дослідницька компанія, проводить опитування, щоб вивчити думки респондентів у багатьох країнах світу щодо реклами в різноманітних медіа, і використовує отриману інформацію для дослідження низки галузей. надається B2B аутсорсингової компанії. 8600 респондентів з таких країн, як: Гонконг, Бразилія, Тайвань, Китай, Австралія, Великобританія, Нідерланди, Іспанія, Австралія, Індія, США. Результати опитування показують, що громадяни готові дивитися більше реклами в Інтернеті та на телебаченні, якщо хочуть отримати фінансову компенсацію. 4 з 10 респондентів відповіли «так». Опитування показало, що 68% вважають, що на телебаченні занадто багато реклами. Інші авіакомпанії також подвоїлися. Іспанія (88%) і Нідерланди (82%) лідирують у розподілі телевізійної реклами, за ними йдуть Бразилія (57%), Тайвань (53%) і Гонконг (38%).

Респонденти в США сказали, що, як правило, занадто багато рекламують як в Інтернеті (56%), так і на радіо (52%). Реклама на мобільних пристроях не залишилася непоміченою, її обсяг різниться залежно від країни і досягає понад 60%. У Бразилії, Австралії, США та Великобританії лише чверть респондентів сказали, що їх дратує мобільна реклама. Таким чином, мобільна реклама також була недоступна в Україні, приблизно 68% респондентів сказали, що вже отримують настирливу та нецікаву рекламу.

Розглянемо також найбільшу рекламну країну та її специфіку. В Америці реклама є синонімом телебачення. У рекламі завжди присутній тиск на споживача, нав'язування, порівняння інших товарів з подібними з метою підкреслення переваг і самого товару Джерело інформації. Тут працює проста, нав'язлива реклама. Актори в 70% рекламних роликів дивляться з екрана прямо на

споживача. Основні цінності, закладені в такій рекламі, - сім'я, любов і патріотизм. Велика увага приділяється слоганам і словам, тому що американці завжди запам'ятовують слогани, які воничують в тій чи іншій рекламі.

Британська реклама абсолютно відрізняється від американської, за рахунок використання гумору в рекламі. Для британців гумор є хорошим способом привернути увагу, сприяти розумінню та покращити настрій і впевненість споживача. Британська реклама завжди має аристократичний акцент, і реклама завжди якісна, ретельно продумана, креативна і з тонким почуттям гумору. Звичайно, в Англії поважають традиції, етикет, манери та хороший смак.

Японська реклама відрізняється від інших унікальним японським поєднанням інновацій і традиційного консерватизму. Така реклама відображає глибокі образи, багату уяву, чудові краєвиди, красиві та яскраві сюжети. Реклама заснована на системі символічного мислення і невербальних знаків, які японці намагаються донести до споживачів. Чуттєве сприйняття залишається важливим для японців у рекламі, оскільки всі вони емоційні. Звичайно, це позитивні емоції під цікавий звуковий супровід.

Французька реклама спрямована на спокушання споживача. Любов, забливість, видовищність і гумор – основні риси реклами.

La Seduction – ця реклама ефективно переконує потенційних споживачів зробити покупку.

Le spectacle – це яскрава реклама з історією та ігровим процесом, що відображає особливості продукту. Французи кажуть, що єдина хороша реклама – це та, яка має відтінок драматизму і носить розважально-театральний характер.

Le Amour – як і у всій Франції, романтика є невід'ємною частиною культури.

Le humour – гумор займає центральне місце в цій рекламі. Однак це не простий гумор, він може приймати різні форми і часто трохи шокує навіть представників Америки.

Німецька реклама – це лише спроба продати товар чи послугу. Німці відомі якістю всіх своїх робіт. Ключові факти, максимальна довіра, почуття

відповідальності, простота викладу та переконливі суворі аргументи є ключовими в цій рекламі.

Українська реклама поширена всюди, і в деяких випадках споживачам вже набридла. Найпоширенішою є телевізійна реклама, за нею йде реклама в Інтернеті. Українська реклама має свою особливість завдяки чітко визначеним вступу, тексту та завершенню. Використовуються популярні треки, смішні картинки, популярні зірки, сленгові фрази та багато іншого. Основними функціями рекламної комунікації є нагадування, психологічна та інформативна. Кожна країна має абсолютно індивідуальні особливості розробки та подання реклами споживачам, що визначаються політичними, соціальними та економічними принципами, традиціями та національністю. Розробляючи оголошення, ви повинні знати про такі особливості та, звичайно, спостерігати за розвитком глобальних оголошень. Абсолютно всі фактори, а саме: особливості соціального, інноваційного та географічного мислення та сприйняття стосуються практично всіх рекламних компаній. Тому ми маємо дуже ретельно працювати з кожною рекламою в кожній країні, щоб ефективно працювати на ринку.

Швидкий розвиток ринку реклами вимагає від його учасників детального вивчення аспектів функціонування, а також зумовлює необхідність швидкого адаптування та прогнозування тенденцій для майбутніх рекламних компаній. Послугами рекламної індустрії користуються суб'єкти усіх видів економічної діяльності, водночас рекламні компанії розвивають кооперацію з суб'єктами інформаційних діяльностей, які надають ефірний час у сфері теле та радіомовлення, з веб-сервісами, з поліграфічними підприємствами, які тиражують друковану рекламну продукцію тощо. При цьому рекламна відповідальність не завжди передається на аутсорсинг спеціалізованим компаніям в рекламній індустрії. Деякі підприємства іноді самостійно розробляють рекламний контент. Ще одна особливість рекламних агентств полягає в тому, що вони мають мультигалузеву приналежність, а саме – до медіаіндустрії, і до креативних індустрій, оскільки кожний рекламний продукт є унікальним результатом творчої

праці та для наявного результату потребує втручання висококваліфікованих спеціалістів.

У періоди посилення конкуренції попит на рекламу зростає. Це важливий елемент у комунікаційній політиці компанії та абсолютно необхідний для того, щоб відрізнити певні продукти від інших, іноді таким чином, щоб вони були більш якісними та конкурентоспроможними. Вони не мають можливості забезпечити достатнє фінансування для просування своїх товарів і послуги. Проте компанії, які не можуть особисто гарантувати поширення інформації про свою продукцію серед людей у великих масштабах, постійно інвестують у різні форми реклами, що за сприятливих обставин може призвести до збільшення обсягів продажів та ринків збуту.

За останні роки рекламний ринок у всьому світі зазнав серйозних змін. По-перше, відбувається зміна його структури. У значній частині ринку реклами та комунікацій незмінно домінує медіакомпонент, а саме реклама в ЗМІ. Дослідження ринку реклами у 2020 році, проведене Всеукраїнською рекламною спілкою спільно з професійними асоціаціями та найбільшими гравцями рекламного ринку, показало, що обсяг реклами на медіаринку становить 24 517 млн грн рисунок 1.1. Немедійна ринкова складова, включаючи маркетингові послуги та ринок PR, сягне 6,252 млн грн. [3].

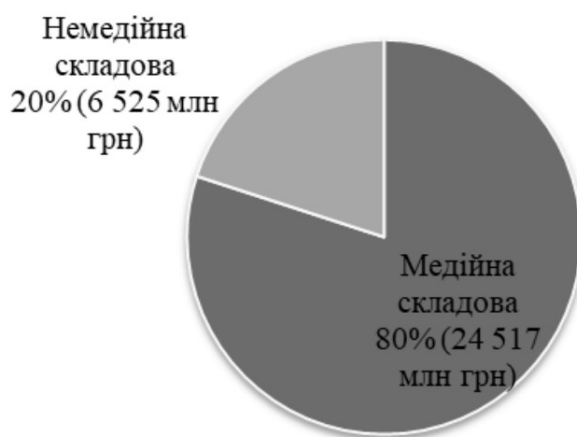


Рисунок 1.1 – Обсяг рекламного ринку України за 2020 рік

Реклама в інтернеті має деякі переваги, а саме: охоплення більшої аудиторії, можливість таргетування, відслідковування зворотного зв'язку, висока швидкість передачі контенту, та його гнучкість, тобто оперативне редагування, доопрацювання, мультимедійність, мультиплатформенність.

Спостерігаються й інші тенденції, окрім збільшення частки медійної реклами в структурі ринку. Зокрема, активними учасниками ринку реклами стають медіаблогери. Рекламодавці співпрацюють з абсолютно різними групами блогерів, наприклад, такими, що мають велику аудиторію або навпаки, невелику аудиторію, але саме вона є цільовою для рекламодавця. Варто зауважити, що не завжди широка аудиторія блогера може гарантувати високу ефективність рекламного звернення. В той же час, розміщення реклами у невеликого блогера з вузько направленою аудиторією зазвичай забезпечує кращу лояльність у потенційного споживача, і в результаті рекламодавець отримує більший дохід, за умови збігу тематики реклами та концепції блогу.

Зацікавлення аудиторій до контенту, який створюють блогери, спираючись на власний досвід, стимулюватиме розвиток існуючих і появу нових платформ для комунікації, а також попит на різні форми контенту. Трендом на сьогодні є збільшення попиту на підкасти, тому велике зацікавлення у споживачів викликала соціальна мережа «Клабхаус», яка більше орієнтована на аудіоконтент, нід якийсь інший. Загалом соціальні мережі мають свої напрямки :

ТікТок– короткі відео, Instagram – фото та відеоматеріали, YouTube – відео, Twitter - короткі повідомлення, цу дозволяє не лише уникнути конфлікту інтересів, але і залучити одних і тих же послідовників у різні соціальні мережі. Вагомим чинником розвитку соціальних мереж є доступ до них з мобільних пристроїв. В Україні Instagram посідає дев'яте місце в списку найбільш відвідуваних веб-сайтів за версією Similar Web, веб-сайту, який надає послуги веб-аналітики для бізнесу.

Для замовлення професійної рекламної продукції компанії звертаються до професійних операторів рекламного ринку. Найвідомішими в Україні є «Fedorig

Agency», «Leo Burnett Ukraine», «Provid» та «Banda Agency». Хоча серед операторів української рекламної індустрії досі немає великих бізнес-одиниць, внесок середніх компаній дуже важливий, згідно з даними Держстату, в обсяг продукції, виробленої в українській промисловості. 47% Це дозволяє зробити висновок про важливість брендів для комерційного успіху також і в рекламному бізнесі. Тому одним із останніх трендів відомих рекламних агентств є розвиток закордонної економічної діяльності. Водночас тенденцією у рекламному бізнесі є зростання його інвестиційної привабливості: збільшення кількості суб'єктів малого підприємництва на 26%, обсяг виробленої ними продукції виріс у 2,7 рази у 2021 р., порівнюючи з 2015 р., і, незважаючи на скорочення на 13% кількості суб'єктів середнього підприємництва, обсяг виробництва у рекламному бізнесі середнього масштабу майже удвічі.

Так як реклама є дуже важливим елементом для ефективної діяльності підприємства (рисунки 1.2, 1.3), її проектуванню та створенню необхідно приділяти неабияке значення. В умовах постійної зміни тенденцій ринку реклами важливим є вміння вчасно пристосуватися до сучасного світу та адаптувати свою діяльність до нових вимог ринку, наприклад, персоналізації рекламних звернень. Змістовна, соціально орієнтована, актуальна та соціально орієнтована реклама завжди буде позитивно впливати на поведінку аудиторії, буде формулювати привабливий образ товару та буде стимулювати скористатися послугами.

Напрямок (згідно із класифікацією IAB)	2015 рік млн грн	2016 рік млн грн	Зміни у % 2016/2015	Прогноз на 2017 рік, млн грн	Зміни у % 2016/2017
Пошук (платна видача в пошукових системах), разом із частиною GDN	800	1200	36%	1650	38%
Банерна реклама, об'яви в соціальних мережах, rich медіа	680	750	10%	830	11%
Спонсорство	120	160	33%	180	13%
Мобільна реклама	150	250	67%	370	48%
Цифрове відео, разом із Youtube	325	460	42%	580	26%
Разом ринок за класифікацією IAB	2155	2820	31%	3610	28%
Інший digital	200	320	60%	400	25%
Разом	2355	3140	33%	4010	28%

Рисунок 1.2 – Обсяг ринку інтернет реклами України 2015 – 2017 роках

	Пошукове просування (SEO)	Контекстна реклама	Банерна реклама	Social Media Marketing	E-mail реклама
Вивести на ринок невідомий товар	-	+/-	+	+	+/-
Отримати швидкий результат	-	+	+/-	+	+/-
Отримати довгостроковий результат	+	-	-	+	+
Таргетувати рекламні повідомлення	+/-	+	+	+/-	+
Організувати повторні продажі	-	+	+	+	+
Обійти рекламні обмеження	+	-	-	+/-	+/-
Заощадити бюджет	+	+/-	-	+/-	+

Рисунок 1.3 – Напрями ефективного використання інструментів інтернет-реклами

На початку 2020 року майже 60% населення світу користувалися Інтернетом. Це 4,54 мільярда, що на 7% більше, ніж у січні 2019 року. Згідно з підрахунками середнього статистичного аналізу Global Overview, одна людина користується мережею 6 годин 43 хвилини, 40% — у нічному режимі. Більше половини свого часу в Інтернеті люди проводять на своїх смартфонах (50,1%). 53% усіх веб-запитів пов'язані зі смартфонами, 44% – з ПК і ноутбуками, а 3% – з планшетами та іншими пристроями.

Процвітання Інтернет-технологій вплинуло на появу нових підприємств і трансформацію існуючих. Не дивно, що Білл Гейтс колись сказав:

«Якщо ваш бізнес не в Інтернеті, він скоро припинить свою діяльність». Такий інтерес до нових технологій матиме значний вплив на розвиток ринку електронної комерції, що позитивно вплине на зростання онлайн закупівель в Україні - 40% у 2019р. Ще оптимістичніша ситуація за кордоном. Згідно зі звітом Electronic Commerce Foundation, 72% європейців здійснюють покупки в Інтернеті, порівняно з 82% у США та 61% у Китаї.

Статистика показує, що з огляду на те, що 75% користувачів купують товари в Інтернеті принаймні раз на місяць, використання Інтернет-реклами в корпоративних маркетингових стратегіях набагато популярніше, ніж реклама на

телебаченні, радіо та друкованих ЗМІ. За оцінками експертів Всеукраїнської рекламної спілки (ВРК), обсяг ринку інтернет-реклами в 2019 році зріс на 40%, сягнувши 14,5 млрд грн (рис. 1.4) [9; 10] :

Інтернет-реклама	Підсумки 2019 р., млн грн.	Частка мобільного трафіку, %	Прогноз зміни на 2020 р., %
Банерна реклама, оголошення у соціальних мережах, rich media	2 915	65	20
Спонсорство	180	60	0
Цифрове відео, зокрема Youtube	2 645	70	30
Разом інтернет-медіа	5 740	–	24
Пошук (платне видавання у пошукових системах), включаючи частину GDN	8 100	70	37
Інший диджитал	676	–	4
Разом інтернет-ринок	14516	–	31

Рисунок 1.4 – Основні показники ринку інтернет-реклам

Проведені нами дослідження ринку інтернет-реклами дозволяють визначити найбільш популярні та перспективні інструменти інтернет-реклами для ведення бізнесу в Україні.

Мобільна інтернет-реклама – це використання мобільних пристроїв (переважно смартфонів) для розміщення реклами. Мобільна інтернет-реклама часто розміщується в мобільних додатках (іграх, програмах, ноутбуках, VPN тощо) і мобільних версіях веб-ресурсів. Незважаючи на те, що частка мобільної реклами в Україні в 2016 році становила менше 10% від загального обсягу цифрового ринку, експерти прогнозують впевнене зростання в цьому напрямку в майбутньому. Експерти підраховали, що мобільна реклама збільшує продажі в середньому на 24%. Близько 30% бюджетів на цифрову рекламу в США та Європі витрачається на цей вид реклами. Крім того, з 2015 року активно використовується технологія торгів у реальному часі (онлайн-реклама), яка розміщує рекламу в реальному часі, розширюючи базу використання мобільної реклами.

Електронна пошта є універсальним інструментом, який дозволяє вирішити основні завдання інтернет-реклами шляхом розсилки адресної рекламної інформації. У той же час, одна з переваг надсилання електронної пошти полягає в тому, що це відносно легко реалізувати. Експерти рекомендують приділяти більше уваги сегментації, щоб зробити вплив вашої реклами більш цілеспрямованим і підвищити ефективність рекламних компаній.

Контекстна реклама - це вид реклами, який показує рекламу користувачам залежно від вмісту, що відображається на веб-сайті. Контекстна реклама працює тільки при використанні методів сегментації. Наприклад, ви можете сегментувати за географією, галуззю, конкретними характеристиками, пріоритетами та категоріями. Після сегментації вам потрібно вибрати ключові слова та ретельно структурувати своїх рекламодавців. Ще одним напрямком контекстної інтернет-реклами є реклама в пошукових системах (Google).

Social Media Marketing (SMM) – реклама в соціальних мережах. Сьогодні соціальні мережі дуже популярні серед користувачів Інтернету. Сьогодні соціальні мережі слід розглядати не лише як майданчики для обміну інформацією та спілкування, а й як інструменти для просування бізнесу. Згідно з проведеним опитуванням, найпопулярнішими соціальними мережами в Україні є Facebook та Instagram. Відеореклама - розміщення відеороликів в Інтернеті. Сьогодні цей напрямок стрімко набирає популярність. Зараз YouTube є другою за величиною соціальною мережею у світі. Сьогодні вся телевізійна реклама також транслюється в Інтернеті, щоб охопити ширшу аудиторію та підвищити ефективність реклами. Інтернет-відеореклама дешевша за телевізійну, що робить її більш привабливою для рекламодавців. Крім того, відстежувати лояльність користувачів до цього виду реклами набагато простіше, оскільки безліч фільтрів дозволяє отримати детальну статистику. Популярності додають цікаві та оригінальні відео. Зростання популярності SmartTV також сприяло поширенню відеореклами.

Рекламні банери – розміщення графічної, анімаційної та текстової інформації в Інтернеті. Сьогодні він набуває популярності в Україні та за кордоном. RTB – набір технологій, які дозволяють рекламодавцям розміщувати на аукціоні рекламні банери, орієнтовані на певну аудиторію (рис. 1.5) :

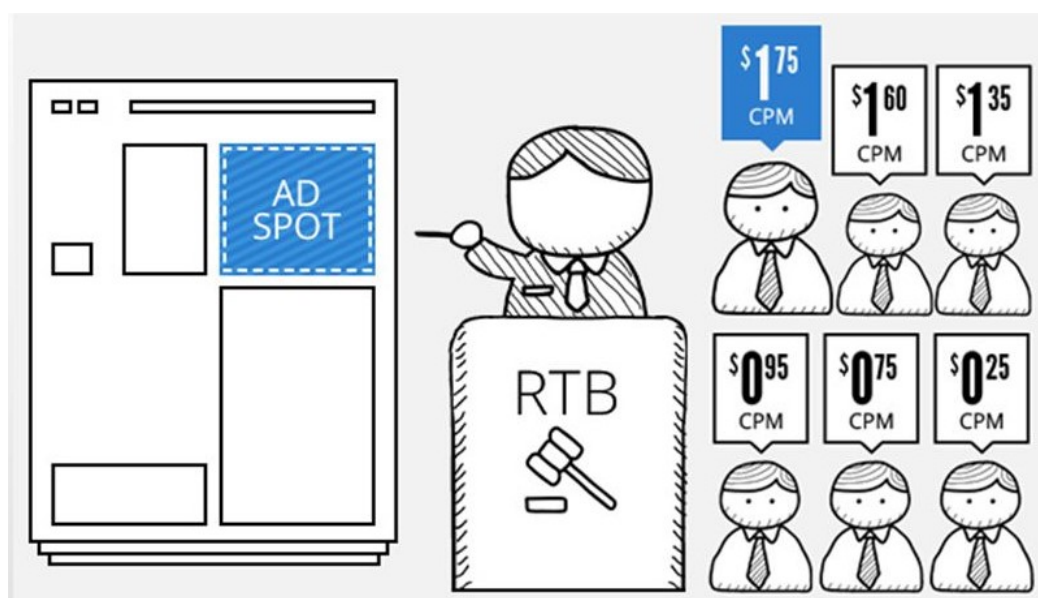


Рисунок 1.5 – Принцип роботи RTB

Так, згідно даних аналізу ринку Інтернет реклами в Україні, можна зробити висновок, що Інтернет-оголошення лишаються перспективним і дієвим інструментом для просування бізнесу. Тенденції розвитку ринку реклами в Україні свідчить про те, що ми поступово переймаємо світовий досвід використання прогресивних Інтернет технологій. Аналіз глобальних тенденцій розвитку Інтернет-технологій, свідчить про те, що в найближчому майбутньому Інтернет-реклама буде все більш ефективною і різноманітною, і її питома вага серед інших видів реклами буде зростати прискореними темпами.

Лідером декілька років є медійна реклама в інтернеті, її частка наблизилася до 40 %, проте вона поступається за популярністю пошуковій рекламі, яка займає 56 % інтернет-ринку. В інтернет-ЗМІ традиційна банерна реклама є найпопулярнішою – 51%, але очікується, що цифрове відео займе верх, з найвищою часткою показів на мобільних пристроях – 70%. Ви можете

використовувати банери та відео для просування власного сайту, інтернет-магазину, бренду, події, а також конкретних продуктів і послуг. Незалежно від того, який тип ви виберете, реклама в Інтернеті може допомогти вам вирішити багато завдань, серед інших:

1. Збільште аудиторію свого веб-сайту, сторінок у соціальних мережах і генерацію потенційних клієнтів.
2. Підвищення впізнаваності бренду.
3. Розширювати нові канали збуту.
4. Знайти нових торгових партнерів (клієнтів, партнерів).
5. Збільшення продажів.
6. Створюйте зворотний зв'язок із клієнтами.
7. Сформууйте прихильність аудиторії до вашого продукту, веб-сайту чи компанії.
8. Інформувати аудиторію про новинки.

Для виконання цих завдань Інтернет переповнений новими рекламними рішеннями, рекламними акціями та інструментами персоналізації. Сьогодні рекламодавці мають доступ до різноманітних оголошень, які найкраще підходять для кожного виду діяльності. Класифікацію Інтернет-реклами вивчали багато вчених.

Дубовик О.В., Бойчук І.В., Л. Ф. Yezhova, О. М. Muzyka, V. V. Тринчук, С. Ю Хамініч. Віртуальна аудиторія висуває настільки високі вимоги до реклами, що набувають популярності нові форми реклами, які витісняють з ринку застарілі рекламні засоби та методи. Рисунок 1.6 враховує останні розробки в Інтернет-секторі щодо вдосконалення класифікації Інтернет-реклами.

Серед користувачів інтернет-продукції побутує думка, що більшість рекламних потоків не є правдивими. Ми провели локальне анкетне опитування

серед студентської молоді та працездатного населення – працівників ЗВО, щоб визначити ставлення користувачів до реклами в інтернеті. У дослідженні взяли участь 63 респонденти у віці 15–50 років. Відповідно до результатів опитування (рис. 1.7).

53 % негативно ставляться до реклами (з них: від 15–30 років – 17 %, від 31–40 років – 28 %, від 41–50 років – 8 %); 35 % – не зважають на неї (з них: від 15–30 років – 8 %, від 31–40 років – 13 %, від 41–50 років – 14 %); 12 % – позитивно ставляться до інтернет-реклами (з них: від 15–30 років – 4 %, від 31–40 років – 2 %, від 41–50 років – 6 %). Респонденти, які входять у вибірку опитаних, вкрай незадоволені агресивними та нав'язливими рекламними банерами pop-up та rich-media, які формують протилежне ставлення до товару/послуги, ніж передбачено метою рекламної кампанії.

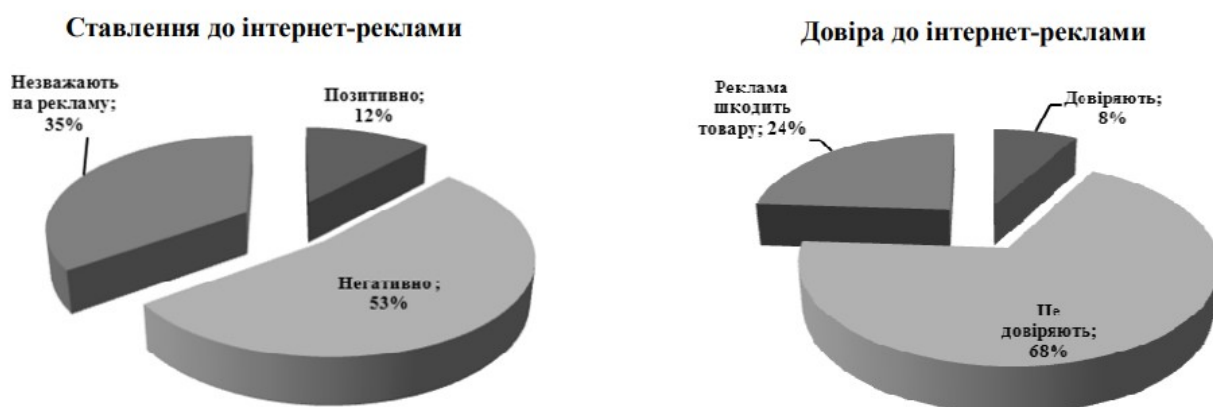


Рисунок 1.6 – Ставлення та довіра інтернет – користувачів до реклами у мережі

Варто відзначити і низький рівень довіри до інтернет-реклами, а саме: довіряють їй – 8 % (із них: 15–30 років – 2 %, 31–40 років – 2 %, 41–50 років – 4 %), не довіряють – 68 % (з них: 15–30 років – 18 %, 31–40 років – 35 %, 41–50 років – 15 %), а 24 % взагалі вважають, що реклама шкодить товару та знижує його рейтинг (з них: 15–30 років – 9 %, 31–40 років – 10 %, 41–50 років – 5 %). Узагальнюючи анкетне дослідження, ми виявили, що найлояльніша до реклами

аудиторія у віці 41– 50 років, у неї найвищий ступінь довіри та позитивне ставлення до реклами.

Незважаючи на упереджене ставлення опитаних респондентів до певних видів реклами, вони продовжують її переглядати та здійснювати покупки завдяки рекламному зверненню, а саму рекламу вважають популярним рішенням просування товарів, послуг. Настрої респондентів щодо інтернет-реклами потрібно враховувати суб'єктам господарювання, особливо на початку рекламної кампанії, коли потрібно визначитися із самим видом реклами, який максимально відповідатиме очікуванням користувачів та меті просування. У зв'язку з наведеною вище інформацією постає потреба виконати порівняльну характеристику видів інтернет-реклами, на які найвищий попит у digital-агентств (рисунок 1.7). Замість того, щоб ігнорувати тип оголошення у своїй рекламній маркетинговій стратегії, варто експериментувати з оголошеннями. SEO-оптимізація допомагає вашому веб-сайту підвищувати рейтинг і з'являтися вище в пошукових запитах. Оптимізація цих позицій займає багато часу, оскільки вся справа в якісному контенті та системному наповненні сторінок. Ефективність SEO-оптимізації спостерігається набагато довше, ніж контекстної реклами. Користувачі Інтернету більше довіряють SEO. Хоча доставка електронною поштою часто дратує користувачів, рекламодавці не відкидають цей спосіб реклами. При цьому керуються його головною перевагою - безкоштовно. Таргетована реклама допомагає оптимізувати продажі та залучити клієнтів через соціальні мережі. Можна налаштувати контекстну та таргетовану рекламу з бюджетом 1 долар на день, але охоплення вашої аудиторії мінімальне та малоефективне. Його ефективність зростає зі збільшенням рекламного бюджету та часу показу.

SEO оптимізація, розсилка електронних листів - самостійно налаштована безкоштовна форма реклами.

Сьогодні ринок Інтернет переповнений великою кількістю рекламних агентств, провайдерів рекламних послуг. Вартість таких послуг залежить від типу

реклами, тривалості, охоплення та географічних особливостей. На рисунку 1.8 наведені результати аналізу цінової політики рекламних агентств. Тому найдешевша для рекламодавців електронна пошта — від 1500 грн на місяць. Вартість контекстної та таргетованої реклами в різних агентствах сильно відрізняється, але ціни майже однакові, починаючи від 5000 грн на місяць.

Вид реклами	Показники				
	Основна мета	Аудиторія	Спосіб оплати	Тривалість отримання рекламного ефекту	Основні показники результативності
Контекстна	прямі продажі	цільова	за кількість переходів на сайт, покази, взаємодію	короткостроковий період	визначається рейтингом кліків (CTR- переходи за посиланням), показник ROI (відношення отриманих вигод до витрат на рекламу)
Таргетингова	інформування, упізнаваність, прямі продажі	цільова	за кількість показів, кліків, переходів, переглядів відео	середньо- та довгостроковий період	покази, частота показів, рекламний бюджет, лід, ціна за лід, CTR, індекс конверсії (кількість продажів/ кількість показів)
SEO-оптимізація (пошукова)	прямі продажі	цільова	безкоштовна	довгостроковий період	рівень підняття вебсторінки у браузері у разі введення ключових слів
E-mail розсилка	інформування, прямі продажі	цільова	безкоштовна	довгостроковий період, рідше коротко- та середньостроковий	індекс конверсії (кількість продажів/ кількість розсилок)

Рисунок 1.7 – Порівняльна характеристика видів інтернет – реклами

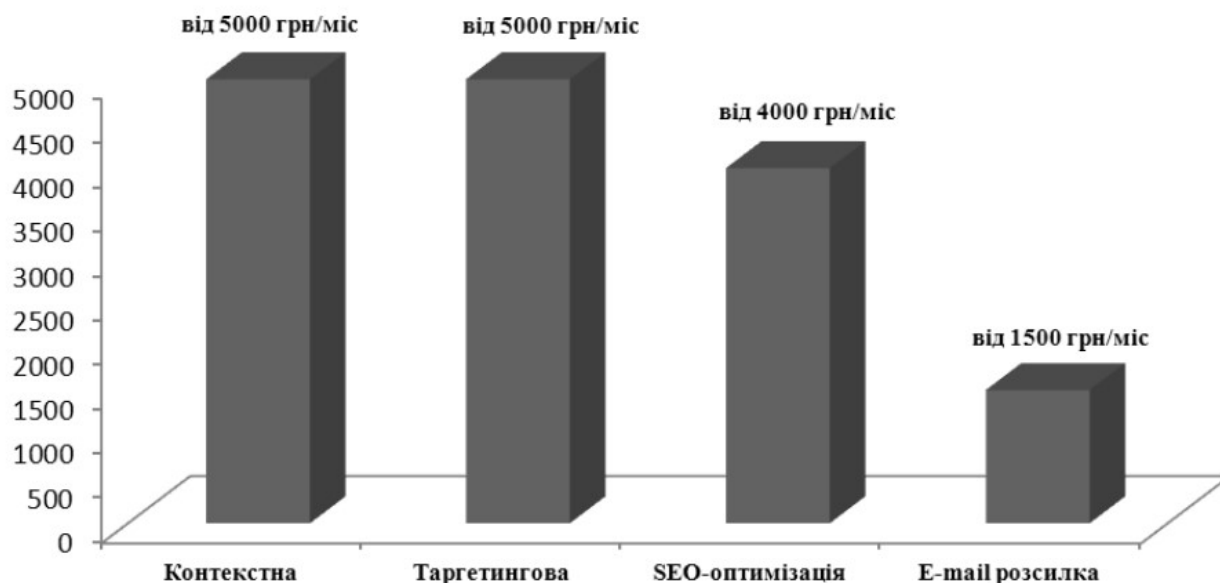


Рисунок 1.8 – Вартість реклами у рекламних агентствах

1.2 Таргетинг – як механізм виділення цільової аудиторії

Цільова аудиторія – це потенційні клієнти, група людей, яка визначається по багатьом параметрам: вік, стать, наявність сім'ї, бюджет сім'ї, соціальний статус, вподобання, розташування, освіта. Рентабельність реклами, яка адаптована під клієнтів має значно більшу вживаність, ніж загальна. На рисунку 1.9 зображено кроки реалізації таргету:

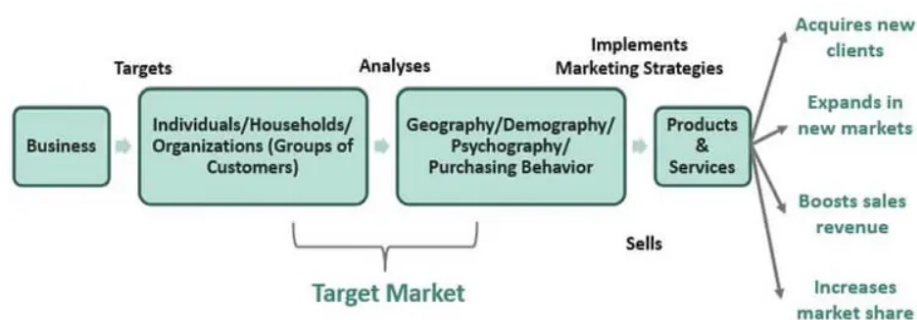


Рисунок 1.9 – Використання таргету

Методи цільової аудиторії:

1. Демографія – соціально-економічні дані, які описують користувача. Ця група включає такі атрибути, як вік, дохід, освіта, стать або географічне розташування.
2. Інтереси – це дані про хобі, пристрасті користувачів, те, що вони шукають і про що читають. Це можуть бути, наприклад, книги, фільми, музика, автомобілі, маркетинг, виховання дітей або танці. Знаючи пристрасті своїх клієнтів, можна отримувати потужні знання, тому що є знання, як залучити клієнтів, який продукт буде для них цікавий. Наприклад, можна запропонувати нову книгу для книголюбів або новий альбом їхньої улюбленої музичної групи.
3. Наміри щодо покупки – дані, які можуть бути вирішальними, особливо для електронної комерції. Аудиторія поділяється на групи користувачів, які нещодавно шукали певний товар, наприклад, ноутбук, холодильник або автомобіль, і раніше його не шукали. Це означає, що вони хочуть купити нову річ, але спочатку потрібно отримати більше знань про затребуваний товар. Прикладами типів аудиторії в цих групах є жіночий одяг, лижі, телевізори, планшети або транспортні засоби. Для націлювання маркетингологи можуть використовувати лише один тип атрибутів аудиторії (наприклад, демографію) або змішувати їх усі, щоб створити детальну цільову групу та точно охопити її.
4. Контекстне націлювання – перший тип націлювання даних – контекстне націлювання. Говорячи про контекстне націлювання, найпростіше подумати про те, що вміст розміщується там, де є ваша аудиторія. У традиційній рекламі контекстне націлювання проявляється в таких способах, як показ вмісту на рекламних щитах у місцях, повз які клієнти проїжджали б, або показ телевізійної реклами під час рекламних пауз у шоу, яке дивився б ідеальний клієнт. Як і слід було очікувати, тепер існують більш просунуті способи націлювання за контекстом, ніж вищезгадані методи. Не потрібно покладатися на рекламні щити чи телебачення в прайм-тайм. Але основна

концепція залишається незмінною. Можна контекстно орієнтуватися на користувачів за допомогою медійної мережі Google. Google зіставлятиме вміст реклами з веб-сайтами, орієнтованими на подібний вміст. Наприклад, це може означати, що оголошення відображатимуться на веб-сайті галузевих новин або на популярному форумі, який зосереджений навколо бізнес-ніші.

5. Фірмографічне націлювання. Фірмографічні дані – це деталі про компанію, за якими вирішується, чи варто на них орієнтуватися. Для B2B-рекламодавців можливість орієнтуватися на основі фірмографічних даних є потужною. Фірмографічні дані дають змогу орієнтуватися на основі таких факторів, як: промисловість, розмір компанії, дохід.
6. Технологічне орієнтування. Націлювання на технічні дані – це спосіб націлювання на облікові записи на основі технологічного пакету компанії. Існує низка дивовижних варіантів використання для технологічного націлювання, і це ідеально, якщо націлюватися на B2B бізнес. Це надає можливість персоналізувати свій вміст і компанії більше, ніж це було б можливо інакше. Незважаючи на зростання популярності протягом останніх років, технічні дані все ще значною мірою недостатньо використовуються багатьма B2B-маркетологами. Загальні заперечення полягають у тому, що це важко отримати або занадто дорого.
7. Націлювання на дані намірів. Доступ до даних про наміри покупців змінює правила гри. Можна використовувати його, щоб дізнатися про компанії в їхньому циклі купівлі, які проблеми вони мають і який тип вмісту вони шукають. Лише одна чверть B2B-компаній використовує дані про наміри у своїх продажах і маркетингу, але з дедалі новими доказами їх ефективності ця кількість зростає. Однією з основних проблем використання даних про наміри є наявність надійного джерела даних. Можна використовувати дані намірів першої сторони, отримані з наявного інструменту аналітики, або отримати доступ до даних намірів за допомогою стороннього

постачальника даних намірів. В ідеалі потрібно використовувати обидва. Це може означати, що використання даних про наміри третьої сторони, щоб відкривати абсолютно нові можливості, наприклад компанії, які шукають рішення. Також потрібно використовувати дані про наміри першої сторони, щоб створити систему оцінки потенційних клієнтів і кваліфікувати потенційних клієнтів, коли вони потрапляють у вашу воронку продажів і маркетингу.

Для того, що визначити свій цільовий ринок, потрібно використати наступні кроки:

1. Переглянути базу попередніх покупок клієнтів, тобто, який товар найбільш вживаний, що клієнтам подобається найбільше у бренді.
2. Виправляти проблеми – добре орієнтуватись у власному продукті, щоб розуміти хто саме зможе та захоче купляти дану послугу.
3. Аналіз продукту – переваги та недоліки бренду.
4. Власна ніша – визначити найунікальнішу функцію бренду.
5. Аналіз конкурентів.
6. Визначити конкретні демографічні групи.
7. Мотивація.

Стратегія планування вибору цільової аудиторії зображена на рисунку 1.10 :

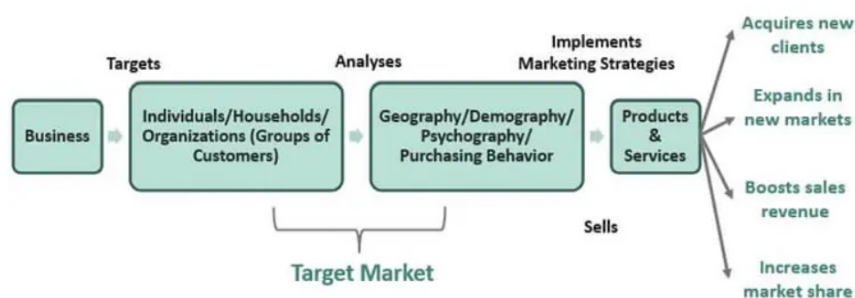


Рисунок 1.10 – Стратегія вибору цільової аудиторії

Для того, щоб бізнес залишався успішним якнайдовше, потрібно розуміти людей, які націлюються. Але перед цим, потрібно пройти три етапи, щоб підтримувати регулярний дохід та розвиток:

1. Сегментація - буде кілька різних типів споживачів, які скористаються послугами. Сегментація дозволяє визначити, хто ці групи, якими, ймовірно, будуть їхні потреби та чи можна їм допомогти. Для окремих компаній і підприємств малого та середнього бізнесу ціна того, щоб бути всім для всіх людей, імовірно, буде високою ціною, тому зосередження на певних сегментах клієнтської бази буде більш вигідним і економічно ефективним. Прагнучи сегментувати свій цільовий ринок, потрібно пам'ятати про змінні, які їх відрізнятимуть, наприклад: місцезнаходження, цінова чутливість, дохід, стать, спосіб життя, цінності, шукані переваги, поведінка. Наприклад, пара, яка щойно народила першу дитину і раніше не думала про страхування життя, може бути більш мотивованою зробити це через новий обов'язок, тоді як успішний бізнесмен на пенсії з великою вартістю майна може розглянути можливість застрахувати життя. Вони обидва хочуть той самий продукт, але причини, чому вони бажають його, абсолютно різні.
2. Націлювання. Незалежно від того, чи вирішено зосередитися на одному чи кількох сегментах, одним із найважливіших елементів, який потрібно визначити, буде те, наскільки добре ці сегменти обслуговуються конкурентами? Чи добре вони відомі? Що вони пропонують? Що ти можна зробити краще? Якщо конкуренти добре обслуговують вибраний сегмент, можливо, варто розглянути, на які інші сегменти можна націлитися. Однак, якщо відомо, що обрана пропозиція має високу цінність диференціації, і можна показати та довести це, тоді варто інвестувати час і ресурси, щоб підвищити впізнаваність бренду.
3. Позиціонування. Отже, визначено сегменти та вирішено, на кого варто орієнтуватися. Останнім кроком є позиціонування бізнесу таким чином,

щоб він відповідав потребам клієнтів і заохочував їх купувати : До якого рівня сприйняття ви прагнете? Ціна – преміум, низька вартість, вартість? Товар – преміум, середній, базовий? Просування по службі – висококласний, доброзичливий і доступний, професійний/консервативний? Розповсюдження – екстенсивне, вибіркоче, ексклюзивне?

Типи націлювання:

1. Поведінкове націлювання – це практика сегментування клієнтів на основі поведінки веб-перегляду, включаючи такі дані, як відвідані сторінки, виконані пошуки, натиснуті посилання та придбані продукти. Якщо додати дані мобільного та фізичного магазину, це також може включати такі речі, як місцезнаходження та покупки в магазині. Потім відвідувачі зі схожою поведінкою групуються у визначені сегменти аудиторії, що дозволяє рекламодавцям націлювати на них конкретні релевантні оголошення та вміст на основі їхньої історії веб-перегляду та покупок. Часто цитованим прикладом поведінкового націлювання є перенацілювання оголошень .
2. Контекстне націлювання передбачає показ оголошень на основі вмісту веб-сайту. Розмістити рекламу посуду на сайті рецептів або рекламу кросівок на форумі для бігу. Це щось на зразок цифрової версії розміщення друкованої реклами в нішевому журналі. Контекстне націлювання працює на основі припущення, що хтось, хто читає сторінку про біг, імовірно, також зацікавиться вашою рекламою кросівок.
3. Перенацілювання на пошук – це коли йде показ медійної реклами користувачам під час перегляду веб-сторінок на основі їх пошукової поведінки за ключовими словами. Кампанії налаштовуються з ключовими словами, які обираються та які відповідають бізнесу чи продуктам. Наприклад, якщо роздрібний продавець меблів, можна показувати медійну рекламу користувачам, які шукали «шкіряний диван» або «шкіряні секції». Такий вид реклами є успішним, тому що він використовує намір зв'язатися

з покупцями. Покупець може знати або не знати про вас, але він виявляє інтерес до продукту чи рішення, які пропонуються.

4. Ретаргетинг сайту. Перенацілювання сайту, передбачає показ медійної реклами користувачам, які відвідали сайт, а потім залишили його, не зробивши покупки, щоб переглянути його в іншому місці. Він відрізняється від пошукового перенацілювання двома важливими моментами: він не базується на ключових словах і націлений на людей, які вже знайомі з вашим брендом, або які принаймні відвідали ваш сайт один раз і виявили інтерес до ваших пропозицій. Завдяки такому визнанню бренду рентабельність інвестицій у ретаргетинг сайту часто надзвичайно висока.
5. Прогнозне націлювання використовує всі дані веб-перегляду за поведінковим націлюванням, шарами сторонніх даних (якщо доступні) і застосовує потужний штучний інтелект і машинне навчання для аналізу даних і прогнозування майбутніх моделей покупок на основі минулої поведінки. Штучний інтелект, який забезпечує прогнозне націлювання, може встановлювати зв'язки між поведінкою, виявляти подібні та споріднені продукти для збільшення та перехресного продажу та зосереджуватися на покупцях, які, найімовірніше, здійснять конверсію в будь-який момент — і все це миттєво. І чим більше даних він аналізує, тим більше він дізнається і тим кращими стають його моделі.

Також для успішної роботи таргету спеціалісти використовують цільові години.

Переваги цільової аудиторії полягають у тому, щоб компаніям ефективніше рекламувати. Хоча важливо охопити якомога більше людей, і часто здається, що зосередження лише на певних сегментах населення є обмеженням, потрібно охопити потенційних споживачів безпосередньо. Припустимо, мета – продати товар працюючим матерям. Кращим підходом до охоплення людей похилого віку

є маркетингова кампанія з використанням друкованих оголошень у газетах і журналах, які мають більш м'який і спокійний тон.

Недоліки таргету:

1. Дорогий і трудомісткий. Компаніям може знадобитися первинне дослідження для вивчення потреб і смаків споживачів. І це коштує чималих грошей. Можливо, компаніям доведеться опитати кількох людей, щоб отримати глибше розуміння. Щоб розробити відповідну маркетингову суміш, учасники також повинні бути залучені в тривалий процес. Компанії повинні вивчати характеристики споживачів, сегментувати їх і вибрати цільові сегменти.
2. Неправильно спрямований. Компанії потребують акуратності в процесі вибору цільових сегментів. Помилки у виборі призводять до невідповідності стратегії та комплексу маркетингу і, зрештою, до провалу продажів. Вибір цільових сегментів також повинен враховувати інші аспекти, а не лише потреби та смаки клієнтів. Компанії повинні враховувати потенціал прибутку сегмента ринку – орієнтація на сегмент, який орієнтується на якість, вигідніша, ніж на сегмент, який орієнтується на ціну. Вони мають відносно високу купівельну спроможність. Розмір сегмента ринку – сегмент ринку має бути достатньо адекватним для отримання ефекту масштабу. Таким чином, компанія може збільшити прибуток разом із збільшенням обсягу продажів. Перспективи зростання – зростаючий сегмент ринку дозволяє компанії продовжувати генерувати продажі. Структурна привабливість – на успіх компанії в цільовому сегменті впливають такі фактори, як інтенсивність конкуренції, переговорна сила покупців і наявність замінників. Усе це впливає на успіх компанії в прибутковому та стабільному продажу продукції. Компанія ігнорує розмір сегмента – через неправильні цілі обсяги продажів низькі і компанія не

може досягти високої економії на масштабі для зниження витрат. Таким чином, прибутків і доходів недостатньо.

1.3 Постановка завдання дослідження

Реклама розвивається, але є такі проблеми як особисті вподобання цільової аудиторії, геополітика, свята та традиції суб'єктів рекламних оголошень, їхні звички та особливості. Тому, коли рекламна організація бере в роботу якийсь проект, їм потрібно провести багато досліджень та аналізу, стосовно аудиторії та самого продукту. Наприклад, якщо потрібно створити рекламу для сервісу техніки Apple, то під час запуску таргетованої реклами рекомендується обирати користувачів, які користуються iPhone, iPad, MacBook, тощо. Також роблячи рекламу для ремонту комп'ютерів потрібно звертати увагу на користувачів, котрі є активними користувачами настільних рішень, замість мобільних.

Для успішної рекламної компанії потрібно купа часу, сил та уваги. При вірному підході рекламної організації до свого замовлення, вони мають створити декілька креативів (відео, фото, музичний супровід) та протестувати кожен набір параметрів (заголовок, креатив, його тривалість та період світового дня), після чого обрати один, або декілька найефективніших та покращувати їх, паралельно знаходячи нові методи та підходи для масштабування рекламної компанії.

Одній людині занадто важко слідкувати одразу за всіма параметрами та враховувати результати показів різних оголошень. Якщо це робить декілька осіб, ситуація не стає менш напруженою. Тому при запуску рекламних компаній, покупці рекламних оголошень користуються певними стереотипами та напрацьованими знаннями, не завжди розбираючи індивідуально кожен продукт.

На ринку є деякі рішення, які самі створюють фото, музику, заголовки та навіть відео, які мають сподобатися користувачеві. Але для аналізу є лише стандартні засоби від Google, які не дають інформацію в зручному для опрацювання вигляді.

Тому було вирішено створити алгоритм для кластеризації суб'єктів реклами, в якому будуть обрані найбільш вагомі параметри для підвищення лояльності цільової аудиторії, що в результаті має привести до збільшення ефективності рекламних компаній.

Мета роботи – оптимізація пошуку цільової аудиторії для рекламної компанії на основі методів машинного навчання.

Об'єкт дослідження – процес кластеризації цільової аудиторії для таргету.

Предмет дослідження – методи машинного навчання.

2 ВИБІР МЕТОДУ МАШИННОГО НАВЧАННЯ ДЛЯ ЕФЕКТИВНОЇ РЕКЛАМИ

2.1 Аналіз методів машинного навчання

Машинне навчання – це галузь дослідження, мета якої є навчити машини виконувати когнітивні завдання так само, як це роблять люди. Хоча вони мають набагато менші когнітивні здібності, ніж звичайна людина, вони здатні швидко обробляти великі обсяги даних і отримувати важливу комерційну інформацію. Алгоритми машинного навчання використовують обчислювальні методи для «вивчення» інформації безпосередньо з даних, а не в залежності від моделі, заснованої на заздалегідь складеному рівнянні. Коли кількість зразків, доступних для навчання, зростає, алгоритми змінюють свою продуктивність.

Групи машинного навчання поділяються на:

1. Контрольоване навчання. Термін «навчання під контролем» відноситься до сценарію, в якому модель використовується для вивчення відображення між вхідними зразками та цільовою змінною. Якщо заздалегідь відомо, чого буде навчатись машина, використовуйте контрольоване навчання. Зазвичай це передбачає піддавання алгоритму великому об'єму навчальних даних, дозволяючи моделі вивчати вихідні дані та тонке налаштування параметрів до отримання бажаних результатів. Потім машину можна протестувати, дозволивши їй генерувати прогнози для «набору даних перевірки» або нових даних, яких раніше не було.
2. Неконтрольоване навчання. Навчання без нагляду дозволяє машині вивчати набір даних без допомоги людини. Після початкового дослідження комп'ютер намагається виявити приховані шаблони, які пов'язують різні змінні. Цей метод навчання може допомогти класифікувати дані за категоріями виключно на основі статистичних атрибутів. Неконтрольоване навчання не потребує великих наборів даних для навчання, що робить його

значно швидшим і легшим у впровадженні, ніж контрольоване навчання. Навчання без контролю, на відміну від навчання під контролем, базується виключно на вхідних даних, без результатів або цільових змінних. Як наслідок, на відміну від навчання під контролем, у навчанні без контролю вчитель не коригує модель.

3. Напівконтрольоване навчання – це контрольоване навчання з невеликою кількістю позначених екземплярів і великою кількістю непозначених прикладів у навчальних даних. На відміну від навчання під контролем, мета моделі напівконтрольованого навчання полягає в тому, щоб ефективно використовувати всі доступні дані, а не лише позначені дані. У напівконтрольованому навчанні використовуються методи навчання без контролю та під контролем. Наприклад, категоризація частини даних вручну може стати прикладом для алгоритму того, як слід сортувати решту набору даних.
4. Навчання з підкріпленням – це техніка, яка дозволяє машині взаємодіяти з оточенням. Простим прикладом є повторна гра у відеогру та винагорода алгоритму, коли він виконує необхідну дію. Зрештою машина може навчитися на своєму досвіді, повторюючи операцію тисячі чи мільйони разів. Модель має певну реакцію, з якої можна навчатися, подібно до навчання під наглядом, хоча зворотний зв'язок може бути затриманим і статистично шумним, що ускладнює для агента або моделі зв'язок між причиною та наслідком.

Алгоритми методів машинного навчання поділяються на ієрархічні та неієрархічні. До перших відносяться: алгоритм середнього сусіда, алгоритм найближчого сусіда, алгоритм дальшого сусіда. До неієрархічних: метод WareCluster, алгоритм Cobweb, алгоритм k-means, алгоритм fuzzy c-means, алгоритм DBSCAN, алгоритм EM. На рисунку 2.1 зображено схему класифікації методів машинного навчання:



Рисунок 2.1 – Методи машинного навчання

Машини градієнтного підвищення — це сімейство потужних методів машинного навчання, які показали значний успіх у широкому діапазоні практичних застосувань. Вони легко налаштовуються відповідно до конкретних потреб програми, наприклад вивчаються щодо різних функцій втрати.

Проблема оцінки функції в класичному навчанні під керівництвом. Той факт, що навчання контролюється, залишає сильне обмеження для дослідника, оскільки дані мають бути забезпечені достатнім набором відповідних цільових міток (вилучення яких може бути дуже дорогим, наприклад, у результаті дорогого експерименту). Мета – реконструювати невідому функціональну залежність так, що деяка задана функція втрат $\Psi(y, f)$ мінімізується (2.1):

$$\hat{f}(x) = y, \quad (2.1)$$

$$\hat{f}(x) = \arg \min_{f(x)} \Psi(y, f(x))$$

На даному етапі не проводиться жодних припущень щодо вигляду ні справжньої функціональної залежності $f(x)$, ні щодо вигляду оцінки функції

$f(x)$. Якщо переписати проблему оцінювання в термінах очікувань, еквівалентним формулюванням буде мінімізація очікуваної функції втрат за змінною відповіді u ($\Psi[u, f(x)]$), залежно від спостережуваних пояснювальних даних x (2.2):

$$\hat{f}(x) = \arg \min_{f(x)} \underbrace{E_x [E_p(\Psi[p, f(x)|x])]}_{\substack{\text{очікувані збитки} \\ \text{очікування для всього набору даних}}} \quad (2.2)$$

Змінна відповіді u може надходити з різних розподілів. Це природно призводить до специфікації різних функцій втрат Ψ . Зокрема, якщо змінна відповіді двійкова, тобто $u \in \{0, 1\}$, можна розглянути біноміальну функцію втрат. Якщо змінна відгуку є безперервною, тобто $u \in \mathbb{R}$, можна використовувати класичну L_2 -квадратичну функцію втрат або надійну регресію Губера. Для інших сімей розподілу відгуків, таких як підрахунок Пуассона, необхідно розробити спеціальні функції втрат. Щоб зробити задачу оцінки функції зрозумілою, потрібно обмежити простір пошуку функції параметричною сім'єю функцій $f(x, \theta)$. Це змінило б задачу оптимізації функції на задачу оцінки параметрів (2.3) (2.4):

$$\hat{f}(x) = f(x, \hat{\theta}) \quad (2.3)$$

$$\hat{\theta} = \arg \min_{\theta} E_x [E_y (\Psi[y, f(x, \theta)]) | x] \quad (2.4)$$

Зазвичай закриті рішення для оцінок параметрів недоступні. Для виконання оцінки розглядаються ітераційні чисельні процедури.

З урахуванням кроків ітерації оцінки параметрів можна записати в інкрементній формі (2.5) :

$$\hat{\theta} = \sum_{i=1}^M \hat{\theta}_i \quad (2.5)$$

Найпростішою і найбільш часто використовуваною процедурою оцінки параметрів є найкрутіший градієнтний спуск. Враховуючи N точок даних (x, y) N $i = 1$ (2.6) :

$$J(\theta) = \sum_{i=1}^N \Psi(y_i, f(x_i, \theta)) \quad (2.6)$$

Класична процедура оптимізації найкрутішого спуску базується на послідовних покращеннях вздовж напрямку градієнта функції втрат $\nabla J(\theta)$. Як оцінює параметр θ^t представлені в інкрементальному порядку, виділити позначення оцінки. За індексом оцінок θ^t ми б розглянули t -й крок приросту оцінки $\hat{\theta}$. Верхній індекс θ^t відповідає згорнутій оцінці всього ансамблю, тобто сумі всіх приростів оцінки від кроку 1 до кроку t . Процедура оптимізації

найкрутішого спуску організована наступним чином:

$$\hat{\theta}^t = \sum_{i=0}^{t-1} \hat{\theta}_i$$

1. Ініціалізація оцінок параметрів θ^0 . Для кожної ітерації t повторити.
2. Отримати складену оцінку параметрів $\hat{\theta}^t$ з усіх попередніх ітерацій (2.7) :

(2.7)

3. Оцінити градієнт функції втрат $\nabla J(\theta)$, враховуючи отримані оцінки параметрів ансамблю (2.8) :

$$\nabla J(\theta) = \{\nabla J(\theta_i)\} = \left[\frac{\partial J(\theta)}{\partial J(\theta_i)} \right]_{\theta=\hat{\theta}^t} \quad (2.8)$$

4. Обчислити нову додаткову оцінку параметра $\hat{\theta}^t$ (2.9) :

$$\hat{\theta}_t \leftarrow -\nabla J(\theta) \quad (2.9)$$

5. Додати новий кошторис $\hat{\theta}^t$ до ансамблю.

Принципова відмінність між методами посилення та звичайними методами машинного навчання полягає в тому, що оптимізація здійснюється у функціональному просторі. Тобто параметризується оцінка функції f в адитивній функціональній формі (2.10) :

$$\hat{f}(x) = \hat{f}^M(x) = \sum_{i=0}^M \hat{f}_i(x) \quad (2.10)$$

У цьому представленні M є кількістю ітерацій, \hat{f}^0 є початковим припущенням $\{\hat{f}_i\}_{i=1}^M$ – це прирости функції, які також називаються «посиленнями». Щоб зробити функціональний підхід можливим на практиці, можна дотримуватися подібної стратегії параметризації сімейства функцій. Тут представлено параметризовані функції $h(x, \theta)$ «базовий навчальний» для того, щоб відрізнити їх від загальних оцінок функції ансамблю $\hat{f}(x)$. Можна вибрати різні сімейства базових модулів навчання, такі як дерева рішень або сплайни.

Різноманітні варіанти моделей базового учня розглядаються та описуються у відповідному розділі цієї статті. Тепер можна сформулювати «жадібний поетапний» підхід збільшення функції з базовими учнями. Для цього на кожній ітерації слід задавати оптимальний розмір кроку ρ . Тому для оцінки функції на t -й ітерації правило оптимізації визначається як (2.11), (2.12):

$$(2.11)$$

$$(\rho_t, \theta_t) = \arg \min_{\rho, \theta} \sum_{i=1}^N \Psi(y_i, \hat{f}_{t-1}) + \rho h(x_i, \theta) \quad (2.12)$$

Для посилення градієнту довільно вказати як функцію втрат, так і моделі базового учня. На практиці, враховуючи деяку конкретну функцію втрат $\Psi(y, f)$ та/або спеціальний базовий навчальний елемент $h(x, \theta)$, розв'язання оцінок параметрів може бути важко отримати. Щоб впоратися з цим, було запропоновано вибрати нову функцію $h(x, \theta_t)$, яка буде найбільш паралельною від'ємному градієнту $\{g_t(x_i)\}_{i=1}^N$ вздовж даних спостереження (2.13):

$$g_t(x) = E_y \left[\frac{\partial \Psi(y, f(x))}{\partial f(x)} \Big| x \right]_{f(x) = \hat{f}_{t-1}(x)} \quad (2.13)$$

Замість того, щоб шукати загальне рішення для збільшення приросту у функціональному просторі, можна просто вибрати новий приріст функції, який буде найбільш корельованим із $-g_t(x)$. Це дозволяє замінити потенційно дуже складну оптимізаційну задачу на класичну мінімізацію за методом найменших

$$\hat{f}_t \leftarrow \hat{f}_{t-1} + \rho_t h(x, \theta_t) \quad \text{квадратів (2.14) :}$$

$$(\rho_t, \theta_t) = \arg \min_{\rho, \theta} \sum_{i=1}^N [-g_t(x_i) + \rho h(x_i, \theta)]^2 \quad (2.14)$$

CART (Дерево класифікації та регресії) є різновидом алгоритму дерева рішень. Він може виконувати завдання як класифікації, так і регресії. Моделі дерева рішень, такі як CART, є життєздатною альтернативою стандартним регресійним моделям і мають іншу спрямованість: результати – за визначенням – безпосередньо спостерігаються та існують групи ризику. CART, як правило, здатний обробляти всі типи факторів дослідження та більшість змінних результатів, забезпечуючи широке поле застосування. Деякі підпрограми CART уже реалізовані в статистичних пакетах, а також доступні як окремі програми.

Дерева класифікації використовуються, коли для кожного спостереження навчальної вибірки відомий клас заздалегідь. Класи в навчальному зразку можуть бути надані користувачем або розраховані відповідно до якогось екзогенного правила. Наприклад, для проекту торгівлі акціями, клас може бути розрахований як предмет реальної зміни ціни активу.

Нехай t_p буде батьківським вузлом, а t_l, t_r - відповідно лівий і правий дочірні вузли батьківського вузла t_p . Нехай вектор класу Y складається з N спостережень із загальною кількістю K класів. Дерево класифікації будується відповідно до правила розбиття - правила, яке виконує поділ навчальної вибірки на менші частини. Вже відомо, що дані кожного разу повинні бути розділені на дві частини з максимальною однорідністю, рисунок 2.2 :

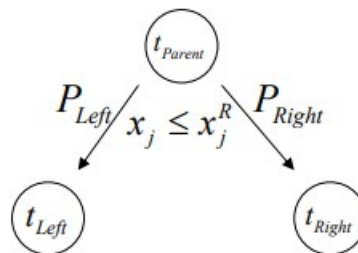


Рисунок 2.2 – Алгоритм розбиття CART

Де t_p , t_l , t_r - батьківський, лівий і правий вузли; x_j - змінна j ; x , P , j – найкраще значення розщеплення змінної x_j . Максимальна однорідність дочірніх вузлів визначається так званою домішковою функцією $i(t)$. Оскільки домішка батьківського вузла t_p постійна для будь-якого з можливих розбиття $x_j \leq x_{Rj}$, $j = 1, M$, буде максимальна однорідність лівих і правих дочірніх вузлів еквівалентна максимізації зміни домішкової функції $\Delta i(t)$ (2.15) :

$$\Delta i(t) = i(t_p) - E[i(t_c)] \quad (2.15)$$

Де t_c - лівий і правий дочірні вузли батьківського вузла t_p . Припускаючи, що P_l , P_r - ймовірності правого і лівого вузлів (2.16) :

$$\Delta i(t) = i(t_p) - P_l i(t_l) - P_r i(t_r) \quad (2.16)$$

Таким чином, на кожному вузлі CART вирішує наступну проблему максимізації (2.17) :

$$\arg \max_{x_j \leq x_j^R, j=1, \dots, M} [i(t_p) - P_l i(t_l) - P_r i(t_r)] \quad (2.17)$$

Рівняння 2.17 передбачає, що CART шукатиме всі можливі значення всіх змінних у матриці X для найкращого розділеного запитання $x_j < x_{Rj}$, що максимізує зміну міри домішки $\Delta i(t)$. Наступне важливе питання полягає в тому, як визначити функцію домішки $i(t)$. Теоретично існує кілька домішкових функцій, але лише дві з них широко використовуються на практиці: правило розбиття Джіні та правило розбиття на дві частини.

Правило розщеплення Джіні (або індекс Джіні) є найбільш широко використовуваним правилом. Він використовує наступну домішкову функція $i(t)$ (2.18) :

$$i(t) = \sum_{k \neq l} p(k|t)p(l|t) \quad (2.18)$$

де $k, l \in 1, K$ - індекс класу; $p(k|t)$ - умовна ймовірність класу k , якщо ми знаходимося у вузлі t . Застосовуючи функцію домішок Джіні 1.18 до задачі максимізації 1.17, буде отримано зміни міри домішки $\Delta i(t)$ (2.19) :

$$\Delta i(t) = - \sum_{k=1}^K p^2(k|t_p) + P_l \sum_{k=1}^K p^2(k|t_l) + P_r \sum_{k=1}^K p^2(k|t_r) \quad (2.19)$$

Таким чином, алгоритм Джіні вирішить наступну задачу (2.20) :

$$\arg \max_{x_j \leq x_j^R, j=1, \dots, M} \left[- \sum_{k=1}^K p^2(k|t_p) + P_l \sum_{k=1}^K p^2(k|t_l) + P_r \sum_{k=1}^K p^2(k|t_r) \right] \quad (2.20)$$

Алгоритм Джіні шукатиме в навчальній вибірці найбільший клас і виділить його решта даних. Сіпні добре працює для шумних даних.

На відміну від правила Джіні, Twoing шукатиме два класи, які складатимуть разом більше, потім 50% даних. Правило поділу на дві максимізує таку зміну міри домішки (2.21) :

$$\Delta i(t) = \frac{P_l P_r}{4} \left[\sum_{k=1}^K \left| p(k|t_l) - p(k|t_r) \right| \right]^2 \quad (2.21)$$

Що передбачає наступну проблему максимізації (2.22) :

(2.22)

Незважаючи на те, що правило розбиття на дві частини дозволяє будувати більш збалансовані дерева, цей алгоритм працює повільніше, ніж правило Джіні. Наприклад, якщо загальна кількість класів дорівнює K , ніж у $2K-1$ можливих поділів. Окрім згаданих правил розбиття Джіні, існує ще кілька методів. Серед найбільш використовуваних є правило ентропії, χ правило, правило максимального відхилення. Але має доведено остаточне дерево нечутливе до вибору правила розбиття. Це процедура, яка є набагато важливішою. Можна порівняти два дерева, побудувати на тому самому наборі даних, але використовуючи різне розбиття правил. За допомогою команди `cartdrawpdfclass` у `XploRe` можна створити дерево класифікації в PDF, вказуючи, яке правило поділу слід використовувати (налаштування третій параметр до 0 для правила Gini або до 1 - для Twoing):

1. `cartdrawpdfclass(x,y, 0, 1)` - Gini splitting rule.

$$\arg \max_{x_j \leq x_j^R, j=1, \dots, M} \left(\frac{P_l P_r}{4} \left[\sum_{k=1}^K |p(k|t_l) - p(k|t_r)| \right]^2 \right)$$

2. `cartdrawpdfclass(x,y, 1, 1)` - Twoing splitting rule.

Дерева регресії не мають класів. Замість цього є вектор відповіді Y , який представляє значення відповіді для кожного спостереження в змінній матриці X . Оскільки регресія дерева не має попередньо призначених класів, правил розбиття класифікації, як Gini або Twoing не можна застосувати. Розбиття в деревах регресії здійснюється відповідно до мінімізації квадратів залишків алгоритм, який

передбачає, що очікувана дисперсія суми для двох результуючих вузлів повинна бути зведена до мінімуму (2.23) :

$$\arg \min_{x_j \leq x_j^R, j=1, \dots, M} [P_l \text{Var}(Y_l) + P_r \text{Var}(Y_r)] \quad (2.23)$$

де $\text{Var}(Y_l)$, $\text{Var}(Y_r)$ – вектори відповіді для відповідних лівого та правого дочірніх вузлів; $x_j \leq x_j^R$, $j = 1, M$ – оптимальне питання розбиття, яке задовольняє умову 1.23. Алгоритм мінімізації квадратичних залишків ідентичний правилу розбиття Джіні. Функцію домішок Джіні 1.19 легко інтерпретувати через позначення дисперсій. Якщо призначати до об'єктів класу k значення 1, а значення 0 об'єктам інших класів, потім вибірка дисперсії цих значень дорівнюватиме $p(k|t)[1 - p(k|t)]$. Підведення підсумків за номером класів K буде отримати таку домішкову міру $i(t)$ (2.24) :

$$i(t) = 1 - \sum_{k=1}^K p^2(k|t) \quad (2.24)$$

До цього моменту було побудовано так зване максимальне дерево, що означає розщеплення було зроблено до останніх спостережень у навчальній вибірці. Максимум дерев може бути дуже великим, особливо у випадку дерев регресії, коли кожне значення відповіді може призвести до окремого вузла.

Переваги CART методу як класифікаційного методу:

1. CART є непараметричним. Тому цей метод не потребує специфікації будь-якої функціональної форми
2. CART не вимагає попереднього вибору змінних. Алгоритм CART самостійно визначить найважливіші змінні та усуне несуттєві. Щоб перевірити цю властивість, можна включити незначну (випадкову) змінну та порівняти нове дерево з деревом, побудованим на початковому наборі

даних. Обидва дерева слід вирощувати з використанням однакових параметрів (правило розбиття та параметр N_{min}).

3. Результати CART інваріантні до монотонних перетворень його незалежних змінних. Зміна однієї або кількох змінних на її логарифм або квадратний корінь не змінить структуру дерева. Лише значення розбиття (але не змінні) у питанні будуть різні.
4. CART може легко обробляти викиди. Викиди можуть негативно вплинути на результати деяких статистичних моделей. Але алгоритм розбиття CART легко впорається з шумовими даними: CART виокремить викиди в окремому місці вузла. Ця властивість дуже важлива, оскільки фінансові дані дуже часто мають викиди через фінансову кризу або дефолт.

Недоліки методу CART:

1. CART може мати нестабільні дерева рішень. Незначна модифікація навчальної вибірки, наприклад виключення кількох спостережень, може призвести до радикальних змін у дереві рішень: збільшення або зменшення дерева, складність, зміни в розподілі змінних і значень.
2. CART розділяє лише одну змінну. Іншими словами, всі розколи перпендикулярні до осі. Зрештою CART виросте величезне дерево, де майже кожне спостережена межа буде в окремому вузлі. Але незважаючи на велике дерево, класифікація буде виконана правильно, і всі спостереження, які належать до червоного класу, будуть класифіковані як червоний, зелений, спостереження як зелений тощо.

Дерева рішень (DTs) — це непараметричний контрольований метод навчання, який використовується для класифікації та регресії. Мета полягає в тому, щоб створити модель, яка передбачає значення цільової змінної шляхом вивчення простих правил прийняття рішень, виведених із характеристик даних. Дерево можна розглядати як кусково-постійне наближення. Наприклад, у

наведеному нижче рисунку 2.3 чим глибше дерево, тим складніші правила прийняття рішень і точніша модель:

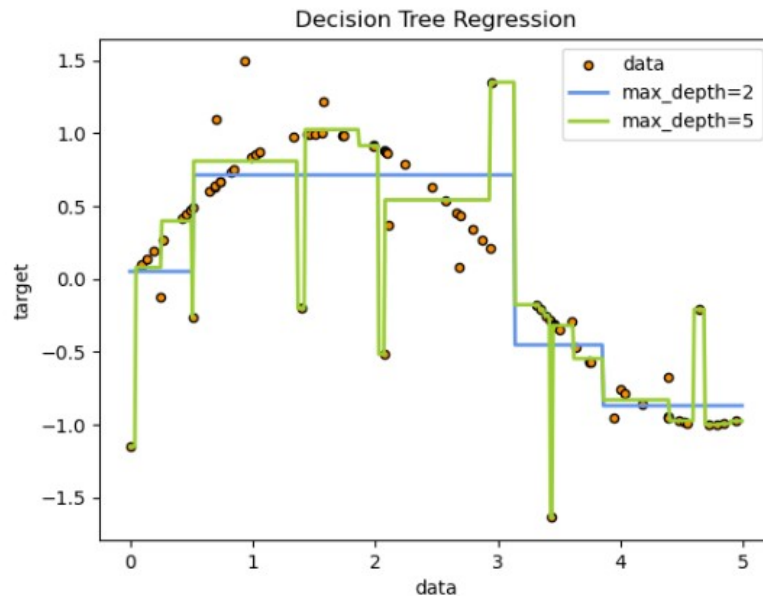


Рисунок 2.3 – Дерево рішень регресії

Переваги дерев рішень:

1. Простий для розуміння та інтерпретації. Дерева можна візуалізувати.
2. Вимагає невеликої підготовки даних. Інші методи часто вимагають нормалізації даних, необхідно створити фіктивні змінні та видалити порожні значення.
3. Вартість використання дерева (тобто прогнозування даних) є логарифмічною в кількості точок даних, які використовуються для навчання дерева.
4. Здатний обробляти як числові, так і категоріальні дані. Однак реалізація `scikit-learn` наразі не підтримує категоріальні змінні. Інші методи зазвичай спеціалізуються на аналізі наборів даних, які мають лише один тип змінної.
5. Здатний вирішувати проблеми з кількома виходами.
6. Використовує модель білого ящика. Якщо дана ситуація є спостережуваною в моделі, пояснення умови легко пояснити булевою логікою. Навпаки, у

моделі чорного ящика (наприклад, у штучній нейронній мережі) результати може бути складніше інтерпретувати.

7. Можливість перевірити модель за допомогою статистичних тестів. Це дозволяє враховувати надійність моделі.
8. Працює добре, навіть якщо його припущення дещо порушуються справжньою моделлю, з якої були згенеровані дані.

До недоліків дерев рішень належать:

1. Щоб уникнути проблеми переобладнання, необхідні такі механізми, як обрізка, встановлення мінімальної кількості зразків, необхідних у вузлі листа, або встановлення максимальної глибини дерева.
2. Дерева рішень можуть бути нестабільними, оскільки невеликі варіації в даних можуть призвести до створення зовсім іншого дерева. Цю проблему можна пом'якшити, використовуючи дерева рішень у ансамблі.
3. Прогнози дерев рішень не є ні гладкими, ні безперервними, а кусково-постійними наближеннями, тому вони погано піддаються екстраполяції.
4. Проблема вивчення оптимального дерева рішень, як відомо, є NP-повною за кількома аспектами оптимальності та навіть для простих понять. Отже, практичні алгоритми навчання дерева рішень базуються на евристичних алгоритмах, таких як жадібний алгоритм, де локально оптимальні рішення приймаються на кожному вузлі. Такі алгоритми не можуть гарантувати повернення глобально оптимального дерева рішень. Це можна пом'якшити, навчивши кілька дерев у груповому модулі, де функції та зразки відбираються випадковим чином із заміною.
5. Існують концепції, які важко вивчити, оскільки дерева рішень нелегко виражають їх, наприклад XOR, проблеми парності або мультиплектора.

6. Учні дерева рішень створюють упереджені дерева, якщо деякі класи домінують. Тому рекомендується збалансувати набір даних перед підгонкою до дерева рішень.

Алгоритми дерева: ID3, C4.5, C5.0 і CART:

1. ID3 (Iterative Dichotomiser 3) був розроблений у 1986 році Россом Квінланом. Алгоритм створює багатостороннє дерево, знаходячи для кожного вузла (тобто в жадібний спосіб) категоріальну ознаку, яка дасть найбільший приріст інформації для категоріальних цілей. Дерева вирощуються до максимального розміру, а потім зазвичай застосовується етап обрізання, щоб покращити здатність дерева узагальнювати невидимі дані.
2. C4.5 є наступником ID3 і усунув обмеження щодо того, що ознаки повинні бути категоричними, динамічно визначаючи дискретний атрибут (на основі числових змінних), який розбиває безперервне значення атрибута на дискретний набір інтервалів. C4.5 перетворює навчені дерева (тобто вихідні дані алгоритму ID3) у набори правил якщо-тоді. Потім оцінюється точність кожного правила, щоб визначити порядок, у якому їх слід застосовувати. Відсікання виконується видаленням передумови правила, якщо без нього точність правила покращується.
3. C5.0 – він використовує менше пам'яті та створює менші набори правил, ніж C4.5, але є точнішим. CART (дерева класифікації та регресії) дуже схожий на C4.5, але відрізняється тим, що він підтримує числові цільові змінні (регресія) і не обчислює набори правил. CART створює двійкові дерева, використовуючи функцію та порогове значення, які дають найбільший приріст інформації на кожному вузлі. scikit-learn використовує оптимізовану версію алгоритму CART; однак реалізація scikit-learn наразі не підтримує категоріальні змінні.

Математичне формування дерева рішень відбувається наступним чином. Задані навчальні вектори, $i=1$ та вектор мітки, дерево рішень рекурсивно розділяє простір функцій таким чином, що зразки з однаковими мітками або схожими цільовими значеннями групуються разом. Нехай дані на вузлі m були представлені як Q із n зразками (2.25) :

$$\begin{aligned} Q_m^{left}(\theta) &= \{(x, y) | x_j \leq t_m\} \\ Q_m^{right}(\theta) &= Q_m \setminus Q_m^{left}(\theta) \end{aligned} \quad (2.25)$$

Якість кандидата на поділ вузла m потім обчислюється за допомогою функції домішок або функції втрат, вибір якого залежить від розв'язуваної задачі (класифікація чи регресія) (2.26) :

$$G(Q_m, \theta) = \frac{n_m^{left}}{n_m} H(Q_m^{left}(\theta)) + \frac{n_m^{right}}{n_m} H(Q_m^{right}(\theta)) \quad (2.26)$$

Далі потрібно обрати домішки (2.27) :

$$\theta^* = \operatorname{argmin}_{\theta} G(Q_m, \theta) \quad (2.27)$$

$$\begin{aligned} Q_m^{left}(\theta) &= \{(x, y) | x_j \leq t_m\} \\ Q_m^{right}(\theta) &= Q_m \setminus Q_m^{left}(\theta) \end{aligned}$$

Критерій класифікації розраховується наступним чином. Якщо ціль це результат класифікації, який приймає значення 0,1, К-1 для вузла m , дозволяє (2.28):

$$p_{mk} = \frac{1}{n_m} \sum_{y \in Q_m} I(y = k) \quad (2.28)$$

Загальні міри домішок Джинні (2.29) :

$$H(Q_m) = \sum_k p_{mk}(1 - p_{mk}) \quad (2.29)$$

Журнал втрати або ентропії (2.1.30) :

$$H(Q_m) = - \sum_k p_{mk} \log(p_{mk}) \quad (2.1.30)$$

$$H(Q_m) = \frac{1}{n_m} \sum_{y \in Q_m} (y \log \frac{y}{\bar{y}_m} - y + \bar{y}_m)$$

$$p_{mk} = \frac{1}{n_m} \sum_{y \in Q_m} I(y = k)$$

Критерії регресії. Якщо ціль є безперервним значенням, то для вузла m загальні критерії, які слід мінімізувати для визначення місць для майбутніх поділів, це середня квадратична помилка (помилка MSE або L2), відхилення Пуассона, а також середня абсолютна похибка (MAE або помилка L1). MSE і відхилення Пуассона встановлюють прогнозоване значення кінцевих вузлів на вивчене середнє значення вузла, тоді як MAE встановлює прогнозоване значення кінцевих вузлів на медіану .

Напівпуассонівське відхилення:

$$\begin{aligned} \text{median}(y)_m &= \text{median}_{y \in Q_m}(y) \\ H(Q_m) &= \frac{1}{n_m} \sum_{y \in Q_m} |y - \text{median}(y)_m| \end{aligned} \quad (2.31)$$

Середня абсолютна похибка:

$$(2.32)$$

Логістична регресія працює дуже подібно до лінійної регресії, але з біноміальною змінною відповіді, може використовувати неперервні пояснювальні змінні та легше обробляти більше двох пояснювальних змінних одночасно. Логістична регресія моделюватиме ймовірність результату на основі індивідуальних особливостей. Оскільки ймовірність — це відношення, фактично моделюватиметься логарифм ймовірності, заданий за формулою (2.33) :

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m \quad (2.33)$$

Де π вказує на ймовірність події (наприклад, смерть у попередньому прикладі) і β_i - це коефіцієнти регресії, пов'язані з групою та x_i – пояснювальні змінні. На цьому важлива концепція повинна бути виділена. Референтна група, представлена β_0 , складається з тих осіб, які представляють еталонний рівень

кожної змінної x_{1m} .

$$H(Q_m) = \frac{1}{n_m} \sum_{y \in Q_m} \left(y \log \frac{y}{\bar{y}_m} - y + \bar{y}_m \right) \quad \text{Модель з усіма включеними}$$

змінними називається «повна модель» або «насичена модель» і є найкращим початковим варіантом, якщо є хороший розмір вибірки та невелика кількість

змінних для включення. Шанси і ймовірності, хоча іноді використовуються як синоніми, не однакові. Ймовірність – це відношення між кількістю подій, сприятливих для деяких результатів і загальна кількість подій. З іншого боку, шанси – це співвідношення між ймовірностями: ймовірність події, сприятливої для результату і ймовірність події проти такого самого результату. Ймовірність обмежена між нулем і одиницею, а шанси обмежені між нулем і нескінченністю. А співвідношення шансів є співвідношенням між коефіцієнтами. Важливість цього полягає в тому, що велике співвідношення шансів (OR) може представляти малу ймовірність і навпаки.

LR передбачає нелінійний зв'язок між залежними та незалежними змінними, а між результатом та значеннями предиктора. Залежна змінна має бути категоричною; незалежні змінні не повинні бути інтервальними; ані нормально розподілені, ані лінійно пов'язані, ані однакові варіації в кожній групі, і, нарешті, категорії (групи) мають бути взаємовиключними та вичерпними. Справа може бути лише в одній групі, і кожна справа має бути членом однієї з груп. LR має можливість враховувати як категоричні, так і безперервні незалежні змінні. Хоча потужність аналізу збільшується, якщо незалежні змінні розподілені нормально і мають лінійний зв'язок із залежною змінною. Перевірка цих припущень показує, що цей метод можна використовувати дещо гнучкіше, ніж традиційні методи регресії, що робить його придатним для багатьох клінічно значущих ситуацій. Для будь-якого заданого випадку LR обчислює ймовірність того, що випадок із певним набором значень для незалежних змінних є членом змодельованої категорії. Потрібні більші вибірки, ніж для лінійної регресії, оскільки максимальні коефіцієнти правдоподібності є оцінками великої вибірки.

Модель логістичної регресії. LR дає кожному предиктору коефіцієнт, який вимірює його незалежний внесок у зміну залежної змінної. Залежна змінна Y приймає значення 1, якщо відповідь «Так», і приймає значення 0, якщо відповідь «Ні». Форма моделі для прогнозованих ймовірностей виражається як натуральний логарифм (\ln) відношення шансів (2.34), (2.35) :

$$\ln \left[\frac{P(Y)}{1-P(Y)} \right] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k \quad (2.34)$$

$$\frac{P(Y)}{1-P(Y)} = e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k}$$

$$P(Y) = e^{\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k} - P(Y) e^{\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k} \quad (2.35)$$

$$P(Y) = \frac{e^{\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k}}{1 + e^{\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k}}$$

де $\ln \left[\frac{P(Y)}{1-P(Y)} \right]$ логарифм (коефіцієнтів) результатів, Y — дихотомічний результат; X_1, X_2, X_k є прогностичними змінними, $\beta_0, \beta_1, \beta_2, \beta_k$ є коефіцієнтами регресії (моделі), а β_0 є перетином. У рівнянні (1.36) модель логістичної регресії безпосередньо пов'язує ймовірність Y зі змінними-прогнозами. Метою LR є оцінка $k + 1$ невідомих параметрів β у рівнянні (1.36). Це робиться за допомогою оцінки максимальної правдоподібності, яка передбачає знаходження набору параметрів, для яких ймовірність спостережуваних даних є найбільшою. Коефіцієнти регресії вказують на ступінь зв'язку між кожною незалежною змінною та результатом. Кожен коефіцієнт представляє величину зміни, яку очікується у змінній відповіді, якби відбулася зміна на одну одиницю в змінній предиктора. Метою LR є правильне прогнозування категорії результату для окремих випадків за допомогою найкращої моделі. Для досягнення цієї мети створюється модель, яка включає всі змінні предиктора, корисні для прогнозування змінної відповіді. LR розраховує ймовірність успіху над ймовірністю невдачі. Результати аналізу представлені у вигляді співвідношення шансів.

Логістична крива. Двійкова залежна змінна має значення 0 і 1, а прогнозоване значення (ймовірність) має бути обмежено, щоб потрапити в той

самий діапазон. Щоб визначити зв'язок, обмежений 0 і 1, LR використовує логістичну криву для представлення зв'язку між незалежною та залежною змінними. За дуже низьких рівнів незалежної змінної ймовірність наближається до 0, але ніколи не досягає 0. Так само, якщо незалежна змінна зростає, прогнозовані значення збільшуються вгору по кривій і наближаються до 1, але ніколи не дорівнюють 1.

Перетворення ймовірності в коефіцієнти й логіт-значення. Логістичне перетворення гарантує, що оцінені значення не виходять за межі діапазону від 0 до 1. Це досягається у два етапи, спочатку ймовірність перераховується як шанси, які визначаються як відношення ймовірності події до ймовірності цього не відбувається. Щоб обмежити прогнозовані значення в межах 0 і 1, значення шансів можна перетворити назад у ймовірність; таким чином (2.36) :

$$Probability(event) = \frac{odds(event)}{1+odds(event)} \quad (2.36)$$

Тому можна показати, що відповідна ймовірність дорівнює $4/(1 + 4) = 0,8$. Крім того, щоб значення шансів були нижчими за 0, що є нижньою межею (верхньої межі немає), необхідно обчислити значення logit, яке обчислюється шляхом логарифмування шансів. Коефіцієнт шансів, менший за 1, має від'ємне логіт-значення, коефіцієнт шансів, більший за 1,0, має позитивне логіт-значення, а коефіцієнт шансів 1,0 (що відповідає ймовірності 0,5) має логіт-значення 0.

Інтерпретація співвідношення шансів (OR). Коли незалежна змінна X_i збільшується на одну одиницю ($X_i + 1$), а всі інші фактори залишаються незмінними, шанси залежної змінної збільшуються на коефіцієнт $\exp(\beta_i)$, який називається відношенням шансів (OR) і діапазонами від нуля (0) до позитивної нескінченності. Він вказує на відносну величину, на яку шанси залежної змінної збільшуються ($OR > 1$) або зменшуються ($OR < 1$), коли значення відповідної незалежної змінної збільшується на одну (1) одиницю.

Вибір залежних змінних. У багатьох випадках цю кінцеву подію легко класифікувати за класами таких, що відбулися чи не відбулися. Як тільки ця категоризація була досягнута, предиктори цього результату можуть бути вивчені. В інших випадках результат можна розглядати як дихотомічний, але насправді він походить від цензури безперервних даних; тобто критерій відсікання був створений, а дані перекодовані з безперервних до категорійних у точці відсікання. У цих випадках ситуація з вибором змінної результату може бути більш складною. У деяких випадках безперервні результати відносно легко перетворюються на дихотомічні події. Ці випадки найчастіше стосуються заходів, для яких були розроблені добре встановлені граничні точки для наявності події. Варто зазначити, що багато багатокатегорійних або навіть безперервних змінних можна звести до дихотомічних.

Вибір потенційних предикторів. Інший аспект, який слід враховувати при розробці дослідження LR, стосується вибору змінних для аналізу як потенційних предикторів результату. Однак існує низка недоліків у виборі змінних предикторів, які можуть призвести до того, що представлена логістична модель, здавалося б, пояснює більшу чи меншу кількість відхилень, ніж вона насправді може пояснити насправді. Результати будь-якого LR залежатимуть від змінних, вибраних як потенційні предиктори, простіше кажучи, якщо змінну не вибрано для аналізу, вона не може бути представлена в остаточній моделі. Однак вибір щодо включення чи ні факторів у вихідний набір даних може вплинути на результати. Крім того, якщо необхідно враховувати умови взаємодії між змінними, то пропуск деяких змінних потенційно може мати серйозний вплив на результати. На жаль, рішення полягає не просто в тому, щоб включити якомога більше змінних, оскільки включення змінних, які не пов'язані з відповідним результатом, це (додавання непов'язаних змінних) має тенденцію завищувати очевидну прогностичну валідність остаточної моделі. Немає жодного найкращого способу визначити, чи обраний набір предикторів є відповідним, але низка практичних правил може показати, що вибір розумний. Наприклад, якщо

специфічність (ступінь, до якої предиктори правильно ідентифікують осіб, які не демонструють конкретний результат, істинно негативні), і чутливість (ступінь, до якої предиктори правильно ідентифікують осіб, які демонструють результат, істинно позитивні) моделі є обидва вище 80%, то ймовірно, що вибрані предиктори мають валідність. Цілком можуть існувати обмеження, що діють на будь-яке конкретне дослідження, що призведе до упередженості у виборі даних, які використовуються для аналізу.

Іншим джерелом упередженості відбору в змінних, які вивчаються, є відсутність даних, де наявність відсутніх даних у вибірці може зменшити розмір вибірки, якщо тих учасників, для яких відсутні дані, виключити, або може призвести до виключення певних змінних з аналізу, якщо відсутні великі обсяги даних. Нарешті, окрім ефектів зміщення відбору з цих джерел, вибір змінних також обмежений властивостями даних, які збираються. Наприклад, змінні предикторів, пов'язані одна з одною (що демонструють колінеарність або мультиколінеарність), або змінні предикторів, які мають надто впливові спостереження (викиди), негативно вплинуть на результати LR. Зокрема, у невеликих або помірних вибірках колінеарність може призвести до загальних рівнів значущості від LR, коли окремі предиктори самі по собі не передбачають результат, або якщо ступінь зв'язку між предиктором і результатом встановлено неправильно. Хоча LR є особливо корисним у забезпеченні економної комбінації найкращих прогностичних змінних, така процедура має тенденцію використовувати випадкові характеристики вибірки. Набір предикторів, отриманий одним зразком, може не відповідати іншому зразку. Тому вважається бажаним під час використання цієї процедури внести поправку на використання випадковості шляхом перехресного повторення на новому зразку.

Відповідність моделі LR можна оцінити кількома способами. По-перше, це оцінити загальну модель (зв'язок між усіма незалежними змінними та залежною змінною). По-друге, необхідно оцінити значущість кожної з незалежних змінних.

По-третє, потрібно оцінити точність прогнозування або дискримінаційну здатність моделі, і, нарешті, модель має бути підтверджена.

Загальна відповідність моделі показує, наскільки сильний зв'язок між усіма незалежними змінними разом і залежною змінною. Його можна оцінити, порівнюючи відповідність двох моделей з незалежними змінними та без них. Кажуть, що модель LR з k незалежними змінними забезпечує кращу відповідність даним, якщо вона демонструє покращення порівняно з моделлю без незалежних змінних (нульова модель). Загальну відповідність моделі з коефіцієнтами k можна перевірити за допомогою тесту співвідношення правдоподібності, який перевіряє нульову гіпотезу (2.37) :

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0 \quad (2.37)$$

Для цього відхилення лише з перехопленням (2 логарифм правдоподібності нульової моделі) порівнюється з відхиленням, коли додано k незалежних змінних (2 логарифм правдоподібності даної моделі). Різниця між ними дає індекс відповідності G . Це показник того, наскільки добре всі незалежні змінні впливають на результат або залежну змінну.

Тест Хосмера-Лемешоу використовується для перевірки того, чи спостережувані частки подій подібні до прогнозованих ймовірностей появи в підгрупах модельної популяції. Тест Хосмера-Лемешоу виконується шляхом розподілу прогнозованих ймовірностей на децилі (10 груп на основі процентних рангів), а потім обчислення χ^2 Пірсона (χ^2), який порівнює прогнозовані частоти з спостережуваними. Значення тестової статистики виражається як (2.38):

$$X = \sum_{g=1}^{10} \frac{O_g - E_g}{E_g} \quad (2.38)$$

Де, O і E позначають спостережувані та очікувані події для g -ї децильної групи ризику. Тестова статистика асимптотично відповідає розподілу χ^2 з 8 (кількість груп мінус 2) DoF. Малі значення (з великим P -значенням, ближчим до 1) вказують на хорошу відповідність даним, отже, хорошу загальну відповідність моделі. Великі значення (з $P < 0,05$) вказують на погану відповідність даним.

Статистична значущість індивідуальних коефіцієнтів регресії. Якщо загальна модель працює добре, наступне питання полягає в тому, наскільки важливою є кожна з незалежних змінних. Коефіцієнт LR для i – i незалежної змінної показує зміну прогнозованих логарифмічних шансів мати результат для однієї одиничної зміни i – i незалежної змінної, за інших рівних умов. Тобто, якщо i – та незалежна змінна з коефіцієнтом регресії b змінюється на 1 одиницю, тоді як усі інші предиктори залишаються постійними, очікується, що логарифмічні шанси результату змінять b одиниць. Існує кілька різних тестів, призначених для оцінки значущості незалежної змінної в логістичній регресії, включаючи тест співвідношення правдоподібності та статистику Вальда.

Статистика Вальда. Статистичні тести значущості можуть бути застосовані до коефіцієнтів кожної змінної. Для кожного коефіцієнта нульова гіпотеза про те, що коефіцієнт дорівнює нулю, перевіряється проти альтернативи того, що коефіцієнт не дорівнює нулю, за допомогою критерію Вальда, V_j . Тест Вальда також можна використовувати для порівняння повної моделі, що містить усі змінні предиктора, зі скороченою моделлю з деякими коефіцієнтами, встановленими на нуль. Статистику Вальда можна використовувати для оцінки внеску окремих предикторів або значущості окремих коефіцієнтів у даній моделі. Статистика Вальда — це відношення квадрата коефіцієнта регресії до квадрата стандартної помилки коефіцієнта. Статистика Вальда асимптотично розподілена як розподіл χ^2 (2.39) :

$$V_j = \frac{\beta_j^2}{SE_{\beta_j}^2} \quad (2.39)$$

Кожна статистика Вальда порівнюється з критичним значенням χ^2 з 1 DoF.

Тест співвідношення правдоподібності, який використовується для оцінки загальної відповідності моделі, також можна використовувати для оцінки внеску окремих предикторів у дану модель. Перевірка співвідношення правдоподібності для певного параметра порівнює ймовірність отримання даних, коли параметр дорівнює нулю, L_0 з ймовірністю L_1 отримання даних, оцінених за максимально правдоподібною оцінкою параметра (2.40) :

$$G = -2 \ln \frac{L_0}{L_1} = -2 \ln (L_0 - L_1) \quad (2.40)$$

Ця статистика порівнюється з розподілом χ^2 з 1 DoF. Щоб оцінити внесок окремих предикторів, можна ввести предиктори ієрархічно, а потім порівняти кожен нову модель з попередньою моделлю, щоб визначити внесок кожного предиктора.

2.2 Вибір та класифікація даних для машинного навчання

Нейронні мережі — це набір алгоритмів, створених за зразком людського мозку, призначених для розпізнавання шаблонів. Вони інтерпретують сенсорні дані за допомогою машинного сприйняття, маркування або кластеризації вихідних даних. Шаблони, які вони розпізнають, є числовими, містяться у векторах, у які мають бути переведені всі дані реального світу, будь то зображення, звук, текст або часові ряди. Нейронні мережі допомагають кластеризувати та класифікувати. Можна розглядати їх як рівень кластеризації та класифікації поверх даних, які зберігаються та керуються. Вони допомагають групувати немічені дані відповідно до подібності серед зразків вхідних даних і класифікують дані, коли мають позначений набір даних для навчання. Нейронні

мережі також можуть витягувати функції, які передаються іншим алгоритмам для кластеризації та класифікації; тому глибокі нейронні мережі розглядаються як компоненти більших програм машинного навчання, що включають алгоритми навчання з підкріпленням, класифікації та регресії.

Усі завдання класифікації залежать від позначених наборів даних, тобто люди повинні передавати свої знання в набір даних, щоб нейронна мережа дізналася про кореляцію між мітками та даними. Це відоме як контрольоване навчання.

1. Розпізнавати обличчя, ідентифікувати людей на зображеннях, розпізнавати вирази обличчя (злість, радість).
2. Визначати об'єкти на зображеннях (знаки зупинки, пішоходи, покажчики смуг).
3. Розпізнавати жести на відео.
4. Виявляти голоси, ідентифікувати мовців, транскрибувати мову в текст, розпізнавати почуття в голосі.
5. Класифікувати текст як спам (в електронних листах) або шахрайський (у страхових претензіях); розпізнавати почуття в тексті (відгуки клієнтів)

Будь-які мітки, які можуть створити люди, будь-які результати, які вас цікавлять і які співвідносяться з даними, можна використовувати для навчання нейронної мережі.

Кластеризація або групування – це виявлення подібності. Для глибокого навчання не потрібні мітки для виявлення подібності. Навчання без міток називається навчанням без нагляду. Дані без міток становлять більшість даних у світі. Один із законів машинного навчання такий: чим більше даних може працювати алгоритм, тим точнішим він буде. Таким чином, неконтрольоване навчання має потенціал для створення високоточних моделей.

1. Пошук: порівняння документів, зображень або звуків із подібними предметами.
2. Виявлення аномалій: зворотною стороною виявлення подібності є виявлення аномалій або незвичної поведінки.

Глибоке навчання — це назва, яка використовується для «стекованих нейронних мереж», тобто мережі, що складаються з кількох шарів. Шари складаються з вузлів. Вузол — це лише місце, де відбуваються обчислення, нещільно побудовані на нейроні в людському мозку, який спрацьовує, коли стикається з достатньою кількістю подразників. Вузол поєднує вхідні дані з набором коефіцієнтів або вагових коефіцієнтів, які або підсилюють, або пом'якшують ці вхідні дані, тим самим призначаючи значення вхідних даних щодо завдання, яке намагається вивчити алгоритм; наприклад, яке введення є найбільш корисним, це класифікація даних без помилок. Ці продукти вхідної ваги підсумовуються, а потім сума передається через так звану функцію активації вузла, щоб визначити, чи має сигнал просуватися далі по мережі, щоб вплинути на кінцевий результат, скажімо, акт класифікації, і в якій мірі. Якщо сигнал проходить, нейрон «активовано», рисунок 2.4 :

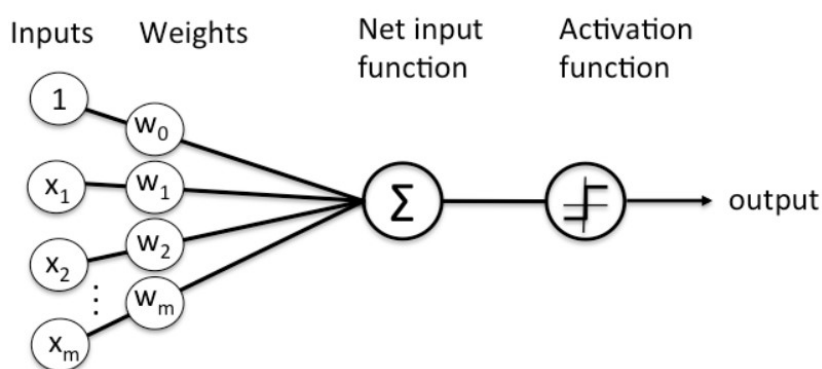


Рисунок 2.4 – Вигляд вигляду елементів нейронної мережі

Нейронні мережі іноді описують з точки зору їх глибини, включаючи кількість шарів, які вони мають між входом і виходом, або так звані приховані

шари моделі. Ось чому термін нейронна мережа використовується майже як синонім глибокого навчання. Їх також можна описати кількістю прихованих вузлів, які має модель, або кількістю входів і виходів, які має кожен вузол. Варіанти класичної конструкції нейронної мережі дозволяють різні форми прямого та зворотного розповсюдження інформації між рівнями.

Типи штучних нейронних мереж включають:

1. Нейронні мережі прямого зв'язку: один із найпростіших варіантів нейронних мереж. Вони передають інформацію в одному напрямку, через різні вхідні вузли, поки вона не потрапить до вихідного вузла. Мережа може мати або не мати прихованих шарів вузлів, що робить їх функціонування більш зрозумілим. Вона готова до обробки великої кількості шуму. Цей тип обчислювальної моделі ШНМ використовується в таких технологіях, як розпізнавання обличчя та комп'ютерне бачення .
2. Рекурентні нейронні мережі більш складні. Вони зберігають вихідні дані вузлів обробки та повертають результат назад у модель. Кажуть, що таким чином модель навчиться передбачати результат шару. Кожен вузол у моделі RNN діє як комірка пам'яті, продовжуючи обчислення та виконання операцій. Ця нейронна мережа починає з того самого фронту поширення, що й мережа прямого зв'язку, але потім продовжує запам'ятовувати всю оброблену інформацію, щоб повторно використовувати її в майбутньому. Якщо передбачення мережі є неправильним, то система самонавчається та продовжує працювати над правильним передбаченням під час зворотного поширення.
3. Згорткові нейронні мережі – одна з найпопулярніших моделей, які використовуються сьогодні. Ця обчислювальна модель нейронної мережі використовує різновид багат шарових персептронів і містить один або більше згорткових шарів, які можуть бути повністю з'єднані. Ці згорткові шари створюють карти функцій, які записують область зображення, яке

зрештою розбивається на прямокутники та надсилається для нелінійного моделювання. Модель CNN особливо популярна у сфері розпізнавання зображень; його використовували в багатьох найсучасніших програмах штучного інтелекту, включаючи розпізнавання облич, оцифрування тексту та обробку природної мови. Інші способи використання включають виявлення перефразування, обробку сигналів і класифікацію зображень.

4. Деконволюційні нейронні мережі : використовують зворотний процес моделі CNN. Вони спрямовані на пошук втрачених функцій або сигналів, які спочатку вважалися неважливими для завдання системи CNN. Цю модель мережі можна використовувати для синтезу та аналізу зображень.
5. Модульні нейронні мережі: містять кілька нейронних мереж, що працюють окремо одна від одної. Мережі не спілкуються та не втручаються в діяльність одна одної під час процесу обчислень. Отже, складні або великі обчислювальні процеси можуть виконуватися більш ефективно.

Переваги штучних нейронних мереж:

1. Можливості паралельної обробки означають, що мережа може виконувати декілька завдань одночасно.
2. Інформація зберігається у всій мережі, а не лише в базі даних.
3. Здатність вивчати та моделювати нелінійні, складні зв'язки допомагає моделювати реальні зв'язки між входом і виходом.
4. Відмовостійкість означає, що пошкодження однієї чи кількох комірок ШНМ не призведе до зупинки генерації виводу.
5. Поступове пошкодження означає, що мережа з часом повільно погіршуватиметься, замість того, щоб проблема миттєво руйнувала мережу.
6. Здатність створювати результати з неповними знаннями з втратою продуктивності залежно від того, наскільки важливою є відсутня інформація.

7. Жодних обмежень не накладено на вхідні змінні, наприклад, як вони мають бути розподілені.
8. Машинне навчання означає, що ШНМ може вчитися на подіях і приймати рішення на основі спостережень.
9. Здатність вивчати приховані зв'язки в даних без встановлення будь-яких фіксованих зв'язків означає, що ШНМ може краще моделювати дуже мінливі дані та непостійну дисперсію.
10. Здатність узагальнювати та виводити невидимі зв'язки з невидимими даними означає, що ШНМ можуть передбачати вихід невидимих даних.

Недоліки штучних нейронних мереж:

1. Відсутність правил для визначення правильної структури мережі означає, що відповідну архітектуру штучної нейронної мережі можна знайти лише шляхом проб, помилок і досвіду.
2. Потреба в процесорах із можливостями паралельної обробки робить нейронні мережі апаратно залежними.
3. Мережа працює з числовою інформацією, тому всі проблеми повинні бути переведені в числові значення, перш ніж їх можна буде представити ШНМ.
4. Відсутність пояснень у рішеннях зондування є одним із найбільших недоліків ШНМ. Неможливість пояснити, чому або як стоїть за рішенням, породжує брак довіри до мережі.

Глибоке навчання використовує нейронні мережі для імітації роботи людського мозку. Тисячі взаємопов'язаних штучних нейронів розташовані в кількох шарах обробки. Два шари загальні з іншими – машинне навчання системи. Додаткові рівні обробки забезпечують абстракції вищого рівня, пропонуючи кращу класифікацію та точніші прогнози. Глибоке навчання ідеально підходить для роботи з великими даними, розпізнавання голосу та спілкування. Часто штучні нейрони мають вагу, яка коригується в міру проходження процесу

навчання. Вага збільшує або зменшує силу сигналу під час підключення. Штучні нейрони можуть мати такий поріг, що сигнал надсилається лише тоді, коли сукупний сигнал перетинає цей поріг. Як правило, штучні нейрони організовані шарами. Різні шари можуть виконувати різні види перетворень на своїх входах. Сигнали проходять від першого (вхідного) шару до останнього (вихідного) шару, можливо, після проходження шарів кілька разів. Алгоритм, який називається вилучення ознак – це ще один аспект глибокого навчання. Цей аспект автоматично створює значущі «функції» з даних для навчання та розуміння.

Алгоритми глибокої нейронної мережі є підрозділ машинного навчання, які використовують «глибоке навчання» для цілей навчання. Ці алгоритми здатні передбачати закономірності, використовуючи попередній досвід. Як форма машинного навчання, глибоке навчання використовує алгоритми для обробки даних і імітації процесу мислення. Глибоке навчання використовує різні рівні алгоритмів для обробки даних, що дозволяє комп'ютеру візуально розпізнавати об'єкти та розуміти людську мову. Дані проходять через кожен шар, а вихід із попереднього шару забезпечує вхід для наступного шару. Найперший рівень мережі називають вхідним. Останній шар називається вихідним. Усі шари між цими двома називаються прихованими шарами. Ці рівні зазвичай використовують прості уніфіковані алгоритми, які містять лише один тип функції активації. Концепція машинного навчання охоплює як робототехніку (робота з реальним світом), так і обробку даних (еквівалент мислення для комп'ютерів). Алгоритми машинного навчання шукають і знаходять передбачувані та повторювані шаблони, які потім можна використовувати для керування даними, електронної комерції та інших нових технологій. Повний вплив машинного навчання тільки починає відчуватися і може суттєво змінити спосіб створення продуктів і те, як люди заробляють на життя. Роботи використовують нейронні мережі, щоб вивчати та передбачати проблеми та шаблони.

2.3 Розробка методики вибору цільової аудиторії

Таргетинг - це механізм, що дозволяє виділити з усіх користувачів лише цільову аудиторію та відображати користувачам той контент, який їм цікавий. Цей інструмент також збільшує користувацьку лояльність. Такі гіганти, як Google та Meta відслідковують буквально кожен ваш крок в інтернеті та знають безліч інформації про кожного користувача, а саме: його IP адресу, країну, місто де він знаходиться, виробник та модель пристрою, версію браузеру, розширення екрану та багато іншого. Володіючи цією інформацією, компанії можуть відображати рекламу вузько сегментованій частині користувачів. Наприклад: показати відео користувачам смартфонів Apple, які знаходяться в Києві та 30км. навколо, ввечері між 20:00 та 24:00. Також є можливість показувати рекламу підписникам свого каналу на YouTube, сторінці Instagram, чи відвідувачам певного сайту. При залученні фахівців можливо відображати відео лише клієнтам, які робили останнє замовлення 3 та більше місяців тому. Крім цього можна додавати безліч фільтрів, в цьому й полягає суть таргетованого маркетингу.

Для успішної рекламної компанії недостатньо обрати геолокацію, додати креатив, обрати стать та вік. Все набагато складніше. Потрібно враховувати звички людей, свята, коли більша частина населення отримує заробітну плату, тестувати різні креативи та враховувати всі показники при наступних запусках реклами. Щоб цього навчитися можна заплатити дуже великі кошти на безліч тестів та навіть не факт, що це обов'язково приведе до бажаних результатів. Саме тому існують фахівці саме в сфері тергету.

Таргетологи використовують різні засоби для аналізу рекламних компаній, деякі з них: Google Analytics, Facebook ADS, Keitaro, Vinom. За допомогою цих засобів фахівці досліджують трафік та змінюють підходи в наступних спробах. Та це завжди займає дуже багато часу. Для вирішення цієї проблеми було вирішено використати методи машинного навчання.

Для досягнення цілі було використано трафік-трекер Keitaro. За допомогою цього інструменту ми отримуємо можливість відслідковувати безліч даних про кожного користувача, який взаємодіє з рекламою. А саме: IP адреса, місто, версія операційної системи та браузеру, розмір екрану, тип підключення та інші. Під час вибору кліків бажано одразу відфільтрувати користувачів, що не є ботами, та обрати лише унікальних. У такому випадку результати будуть більш точними, рисунок 2.5:

Subid	дата і час	IP	ID кампанії	ID потоку	Прапор країни	Регіон	Город	OS	Бри	Тип з'єднани	Тип пристрою	Модель прик	Бот	Уник. (К)	Лід
1a8f6m1md2ko	29 листопада 2022, 00...	99.254.208.190	21	40		Онтаріо	Стрет...	Windows	Chrome	Невідомо	Робочий стіл				✓
2iv8f151md34h	29 листопада 2022, 00...	87.255.94.99	22	41		Кишинів	Киши...	Windows	Chrome	Невідомо	Робочий стіл				
12mka31md4p8	29 листопада 2022, 00...	92.112.60.1	22	41		Ліфанька обл...	Вілово...	Windows	Chrome	Невідомо	Робочий стіл				✓
rqntc1md2kf	29 листопада 2022, 00...	92.112.92.66	22	41		Київ	Київ	Windows	Chrome	Невідомо	Робочий стіл				
2all41md4v4	29 листопада 2022, 00...	60.28.209.8	21	40		Тяньцзінь	Тяньц...	Windows	Chrome	Невідомо	Робочий стіл				✓
1uvq21c1md3i0	29 листопада 2022, 00...	77.87.39.83	22	41		Київ	Київ	Windows	Chrome	Невідомо	Робочий стіл				
1v98j11md2dt	29 листопада 2022, 00...	195.189.80.37	22	41		Варна	Варна	Windows	Chrome	Невідомо	Робочий стіл				
208k93c1md4qu	29 листопада 2022, 00...	92.114.187.77	21	40		Кишинів	Киши...	Windows	Chrome	Невідомо	Робочий стіл				✓
29gmeid1md440	29 листопада 2022, 00...	81.180.65.6	21	40		Кишинів	Киши...	Windows	Chrome	Невідомо	Робочий стіл				✓
4t9gk61md1r1	29 листопада 2022, 00...	77.87.39.83	21	40		Київ	Київ	Windows	Chrome	Невідомо	Робочий стіл				✓

Рисунок 2.5 – Трафік – трекер Keitaro

Саме параметр "Lead" відповідає за успішність рекламної компанії до цього користувача. Значення "1" означає, що цей користувач виконав умови, за якими таргетолог отримав свій прибуток.

Спочатку потрібно перетворити дані у правильний вигляд. Для цього відокремимо годину доби, о котрій користувачі натискали на рекламу, від всієї дати. Далі відобразимо таблицю за допомогою Jupiter, рисунок 2.6 :

The screenshot shows a Jupyter Notebook with the following code and output:

```
In [1]: import pandas as pd
from matplotlib import pyplot as plt

In [2]: PATH_TO_DEFAULT = 'newClicks5.csv'
PATH_TO_TEST = 'newClicks-test.csv'

In [3]: default = pd.read_csv(PATH_TO_DEFAULT) #визначити джерело даних
test = pd.read_csv(PATH_TO_TEST) #джерело даних

In [4]: default
Out[4]:
```

	SubId	Date and time	Hour	IP	Camp. ID	Flow ID	Landing page	Offer	Source	Affiliate Network	Connection type	Device type	Device model	Bot	Unique (C)	UserAgent	Language	Country	Lead	Click ID
0	2ic5m2ndv	02 May 2022 10:39	10	86.49.243.178	72	110	ANNA	Not set	GoogleAds	Not set	WiFi	Mobile phone	Xiaomi Redmi Note 10X	NaN	1	Mozilla/5.0 (Linux; Android 11; M2003J15SC) App...	Czech	Czech Republic	NaN	89283
1	b79r4t2n4	02 May 2022 10:44	10	80.92.246.251	72	110	ANNA	Not set	GoogleAds	Not set	WiFi	Mobile phone	Huawei P30 Lite	NaN	1	Mozilla/5.0 (Linux; Android 10; MAR-LX1A) Appl...	Czech	Czech Republic	NaN	89288
2	27d0rd2nab	02 May 2022 10:57	10	109.105.39.3	72	110	ANNA	Not set	GoogleAds	Not set	WiFi	Mobile phone	Samsung Galaxy S8	NaN	1	Mozilla/5.0 (Linux; Android 9; SM-G950F) Appa...	Czech	Czech Republic	NaN	89275
3	4uc9r2nrf	02 May 2022 12:28	12	193.165.96.252	72	110	ANNA	Not set	GoogleAds	Not set	WiFi	Mobile phone	Xiaomi Redmi 9C NFC	NaN	1	Mozilla/5.0 (Linux; Android 10; M2009C3MNO) Ag...	Czech	Czech Republic	NaN	89343
4	1hceek2nd5	02 May 2022 12:51	12	86.49.80.243	72	110	ANNA	Not set	GoogleAds	Not set	WiFi	Mobile phone	Xiaomi Redmi Note 8 Pro	NaN	1	Mozilla/5.0 (Linux; Android 11; Redmi Note 8 P...	Czech	Czech Republic	NaN	89365
...
2711	29nd09v3pj	31 May 2022 14:29	14	195.245.200.178	78	122	valisa	Not set	GoogleAds	Not set	Unknown	Desktop computer	NaN	NaN	1	Mozilla/5.0 (Windows NT 10.0; Win64; x64) Appl...	Czech	Czech Republic	NaN	118435
2712	29107us3q4	31 May 2022 14:25	14	90.183.80.168	78	122	valisa	Not set	GoogleAds	Not set	Unknown	Desktop computer	NaN	NaN	1	Mozilla/5.0 (Windows NT 10.0; Win64; x64) Appl...	Czech	Czech Republic	NaN	118452
2713	19k2v973j2	31 May 2022 14:48	14	188.120.193.66	78	122	valisa	Not set	GoogleAds	Not set	Unknown	Desktop computer	NaN	NaN	1	Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:1...	Czech	Czech Republic	NaN	118482
2714	302u6t73e1	31 May 2022 14:55	14	194.36.94.181	78	122	valisa	Not set	GoogleAds	Not set	Unknown	Desktop computer	NaN	NaN	1	Mozilla/5.0 (Windows NT 6.1; Win64; x64; rv:10...	Czech	Czech Republic	NaN	118513
2715	3em6283bv	31 May 2022 15:17	15	94.113.125.248	78	122	valisa	Not set	GoogleAds	Not set	Unknown	Desktop computer	NaN	NaN	1	Mozilla/5.0 (Windows NT 10.0; Win64; x64) Appl...	Czech	Czech Republic	NaN	118575

2716 rows x 25 columns

Рисунок 2.6 – Jupyter

Далі перевіримо наші дані на цілісність. Для цього потрібно відобразити інформацію про кожну з колонок таблиці та, якщо деяких даних не вистачає, то потрібно або видалити ці дані, або заповнити. Заповнюють недостатні дані модою, чи середнім арифметичним, у випадку, коли дані числові. Після перевірки отримуємо результат, рисунок 2.7 :

```
In [6]: default.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2716 entries, 0 to 2715
Data columns (total 25 columns):
#   Column                Non-Null Count  Dtype
---  ---                -
0   SubId                 2716 non-null  object
1   Date and time         2716 non-null  object
2   Hour                  2716 non-null  int64
3   IP                    2716 non-null  object
4   Camp. ID              2716 non-null  int64
5   Flow ID               2716 non-null  int64
6   Landing page          2716 non-null  object
7   Offer                 2716 non-null  object
8   Source                2716 non-null  object
9   Affiliate Network     2716 non-null  object
10  Country flag          2716 non-null  object
11  State/region          2716 non-null  object
12  City                  2716 non-null  object
13  OS                    2716 non-null  object
14  Browser               2713 non-null  object
15  Connection type       2716 non-null  object
16  Device type           2716 non-null  object
17  Device model          1828 non-null  object
18  Bot                   0 non-null     float64
19  Unique (C)            2716 non-null  int64
20  UserAgent             2716 non-null  object
21  Language              2716 non-null  object
22  Country               2716 non-null  object
23  Lead                  185 non-null   float64
24  Click ID              2716 non-null  int64
dtypes: float64(2), int64(5), object(18)
memory usage: 530.6+ KB
```

Рисунок 2.7 – Результат заповнених даних у таблиці

Ми бачимо список параметрів та повноту їхнього заповнення в таблиці. В даному випадку ми бачимо, що параметри "Lead", "Device Model", "Browser" заповнені не повністю. Спочатку очистимо вибірку від зайвих даних, які не знадобляться в дослідженні. Враховуючи досвід таргетологів та умови завдання було вирішено виключити такі характеристики, як: "Country", "Browser", "State/region", "Language", "Device model", "Country flag", "Unique (C)", "Bot", "UserAgent", "Date and time", "SubId", "IP", "Offer", "Camp. ID", "Flow ID", "Source", "Affiliate Network", "Connection type", "Click ID", рисунок 2.8 :

```
In [7]: useless_columns = ['Country', 'Browser', 'State/region', 'Language', 'Device model', 'Country flag', 'Unique (C)', 'Bot', 'UserAgent', 'Date and time', 'SubId', 'IP', 'Offer', 'Camp. ID', 'Flow ID', 'Source', 'Affiliate Network', 'Connect
default_clean = default.drop(columns=useless_columns)
useless_columns = ['Country', 'Browser', 'State/region', 'Language', 'Device model', 'Country flag', 'Unique (C)', 'Bot', 'UserAgent', 'Date and time', 'SubId', 'IP', 'Offer', 'Camp. ID', 'Flow ID', 'Source', 'Affiliate Network', 'Connect
test_clean = test.drop(columns=useless_columns)
```

Рисунок 2.8 – Додаткові характеристики

В результаті отримуємо вибірку з такими параметрами: "Hour", "Landing page", "City", "OS", "Device type", "Lead", рисунок 2.9 :

```
In [8]: default_clean
```

```
Out[8]:
```

	Hour	Landing page	City	OS	Device type	Lead
0	10	ANNA	Brno	Android	Mobile phone	NaN
1	10	ANNA	Prague	Android	Mobile phone	NaN
2	10	ANNA	Brno	Android	Mobile phone	NaN
3	12	ANNA	Ostrava	Android	Mobile phone	NaN
4	12	ANNA	Ostrava	Android	Mobile phone	NaN
...
2711	14	valisa	Prague	Windows	Desktop computer	NaN
2712	14	valisa	Prague	Windows	Desktop computer	NaN
2713	14	valisa	Prague	Windows	Desktop computer	NaN
2714	14	valisa	Prague	Windows	Desktop computer	NaN
2715	15	valisa	Prague	Windows	Desktop computer	NaN

Рисунок 2.9 – Вибірка даних

Заповнюємо пропущені значення для категоріальних змінних найбільш частим значенням в колонці, кількісних змінних медіаною по стовпцю, а також замінюємо всі входження значення NaN на нулі у стовпці "Lead", рисунок 2.10 :


```

In [20]: def impute_data(df):
         df_copy = df.copy() #копіюємо, щоб не змінювати оригінал
         #заповнюємо пропущені значення для категоріальних змінних найбільш частим значенням в колонці
         categorial_vars = list(df_copy.select_dtypes(include=['object']).columns.values)
         for col in categorial_vars:
             print(col)
             if (df_copy[col].isnull().values.any()):
                 df_copy[col].fillna(df_copy[col].mode()[0], inplace=True)
         #заповнюємо пропущені значення для кількості змінних медіаною по стовпцю
         numerical_vars = list(df_copy.select_dtypes(include=['int64', 'float64']).columns.values)
         for col in numerical_vars:
             if (col=='Lead'):
                 if (df_copy[col].isnull().values.any()):
                     df_copy[col].fillna(0, inplace=True)
                 continue
             if (df_copy[col].isnull().values.any()):
                 df_copy[col].fillna(df_copy[col].median(), inplace=True)
         return df_copy

In [21]: default_full = impute_data(default_clean)
         test_full = impute_data(test_clean)

```

Рисунок 2.10 – Заповнені значення у таблиці

Отримуємо таблицю такого вигляду, рисунок 2.11 :

```

In [22]: default_full
Out[22]:

```

	Hour	Landing page	City	OS	Device type	Lead
0	10	ANNA	Brno	Android	Mobile phone	0.0
1	10	ANNA	Prague	Android	Mobile phone	0.0
2	10	ANNA	Brno	Android	Mobile phone	0.0
3	12	ANNA	Ostrava	Android	Mobile phone	0.0
4	12	ANNA	Ostrava	Android	Mobile phone	0.0
...
2711	14	valisa	Prague	Windows	Desktop computer	0.0
2712	14	valisa	Prague	Windows	Desktop computer	0.0
2713	14	valisa	Prague	Windows	Desktop computer	0.0
2714	14	valisa	Prague	Windows	Desktop computer	0.0
2715	15	valisa	Prague	Windows	Desktop computer	0.0

Рисунок 2.11 – Готова таблиця даних

Після перевірки на цілісність отримуємо позитивний результат, рисунок 2.12:

```
In [23]: default_full.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2716 entries, 0 to 2715
Data columns (total 6 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Hour             2716 non-null   int64
1   Landing page     2716 non-null   object
2   City             2716 non-null   object
3   OS               2716 non-null   object
4   Device type     2716 non-null   object
5   Lead            2716 non-null   float64
dtypes: float64(1), int64(1), object(4)
memory usage: 127.4+ KB
```

Рисунок 2.12 – Результат заповнених даних

Імпортуємо клас класифікатор `DecisionTreeClassifier` із бібліотеки `sklearn.tree`, рисунок 2.13:

```
In [16]: from sklearn.tree import DecisionTreeClassifier
```

Рисунок 2.13 – Імпорт класифікатора `DecisionTreeClassifier`

Створюємо екземпляр класу `DecisionTreeClassifier`, рисунок 2.14:

```
In [17]: #створення екземпляру класу
dt_clf = DecisionTreeClassifier()

In [18]: y_default_full = default_full['Lead']
default_full_sep = default_full.drop(columns=['Lead'])
```

Рисунок 2.14 – Екземпляр класу `DecisionTreeClassifier`

Клас `DecisionTreeClassifier` має такі параметри: **crit**erion: {"gini", "entropy", "log_loss"}, default="gini".

Функція вимірювання якості розбиття. Підтримувані критерії: «джині» для домішки Джіні та «log_loss» і «ентропія» : **split**ter: {"best", "random"}, default="best".

Стратегія, яка використовується для вибору поділу на кожному вузлі. Підтримувані стратегії: «найкраща» для вибору найкращого розподілу та «випадкова» для вибору найкращого випадкового поділу : **max_depth: int, default=None.**

Максимальна глибина дерева. Якщо немає, то вузли розгортаються, доки всі листи не стануть чистими або поки всі листи не міститимуть менше зразків **min_samples_split: min_samples_split: int or float, default=2.**

Мінімальна кількість зразків, необхідних для розбиття внутрішнього вузла:

1. Якщо int, то вважати min_samples_split мінімальним числом.
2. Якщо float, то min_samples_split є дробом і є мінімальною кількістю вибірок для кожного розбиття $\text{ceil}(\text{min_samples_split} * \text{n_samples})$.

min_samples_leaf: int or float, default=1.

Мінімальна кількість зразків, необхідна для листкового вузла. Точка розділення на будь-якій глибині буде розглядатися, лише якщо вона залишає принаймні min_samples_leaf навчальні зразки в кожній з лівої та правої гілок. Це може мати ефект згладжування моделі, особливо в регресії.

1. Якщо int, то вважати min_samples_leaf мінімальним числом.
2. Якщо float, то min_samples_leaf є дробом і є мінімальною кількістю зразків для кожного вузла. $\text{ceil}(\text{min_samples_leaf} * \text{n_samples})$

min_weight_fraction_leaf: float, default=0.0 – Мінімальна зважена частка загальної суми ваг (усіх вхідних вибірок), яка повинна бути на листовому вузлі. Зразки мають однакову вагу, якщо sample_weight не надано.

max_features: int, float or {"auto", "sqrt", "log2"}, default=None – Кількість функцій, які слід враховувати під час пошуку найкращого розділення:

1. Якщо `int`, тоді враховуйте `max_features` особливості при кожному розбитті.
2. Якщо `float`, то `max_features` є дробом, і функції враховуються при кожному розділенні `max(1, int(max_features * n_features_in_))`.
3. Якщо «авто», то `max_features=sqrt(n_features)`.
4. Якщо «sqrt», то `max_features=sqrt(n_features)`.
5. Якщо «log2», то `max_features=log2(n_features)`.
6. Якщо немає, то `max_features=n_features`.

Примітка: пошук розбиття не припиняється, доки не буде знайдено принаймні один дійсний розділ зразків вузлів, навіть якщо для цього потрібно ефективно перевірити більше ніж `max_features` функції.

`random_state`: *int, RandomState instance or None, default=None* – Контролює випадковість оцінювача. Функції завжди випадково переставляють під час кожного розділення, навіть якщо `splitter` встановлено значення "best". Коли , алгоритм буде вибирати випадковим чином для кожного розбиття перед тим, як знайти найкраще розбиття серед них. Але найкращий знайдений розподіл може відрізнитися в різних прогонах, навіть якщо . Це має місце, якщо покращення критерію є ідентичним для кількох розбивок і один розбиття має бути обрано випадковим чином. Щоб отримати детерміновану поведінку під час підгонки, має бути зафіксовано ціле число.

`max_leaf_nodes`: *int, default=None* – Виростить дерево `max_leaf_nodes` в режимі найкраще. Найкращі вузли визначаються як відносне зменшення домішок. Якщо немає, то необмежену кількість листових вузлів.

`min_impurity_decrease`: *float, default=0.0* – Вузол буде розщеплено, якщо це розбиття спричинить зменшення домішки більше або дорівнює цьому значенню.

Зважене рівняння зменшення домішок має такий вигляд, рисунок 2.15 :

$$N_t / N * (\text{impurity} - N_t_R / N_t * \text{right_impurity} - N_t_L / N_t * \text{left_impurity})$$

Рисунок 2.15 – Зважене рівняння зменшення домішок

де N – загальна кількість вибірок, N_t – кількість вибірок у поточному вузлі, N_t_L – кількість вибірок у лівій дочірній частині та N_t_R – кількість вибірок у правій дочірній частині.

N , N_t , N_t_R , N_t_L всі стосуються зваженої суми, якщо `sample_weight` передано.

class_weight: *dict, list of dict or “balanced”, default=None* - Ваги, пов'язані з класами у формі `dict`. Якщо немає, всі класи повинні мати вагу одиниці. Для проблем з декількома виводами можна надати список диктовок у тому ж порядку, що й стовпці у `{class_label: weight}`.

Зауважте, що для кількох виходів (включно з кількома мітками) ваги повинні бути визначені для кожного класу кожного стовпця у власному `dict`. Наприклад, для класифікації з чотирма класами з кількома мітками ваги мають бути `[[0: 1, 1: 1], {0: 1, 1: 5}, {0: 1, 1: 1}, {0: 1, 1: 1}]` замість `[[1:1], {2:5}, {3:1}, {4:1}]`.

«Збалансований» режим використовує значення `u` для автоматичного коригування ваг обернено пропорційно частотам класів у вхідних даних як `n_samples / (n_classes * np.bincount(y))`

Для мультивиходу ваги кожного стовпця у будуть перемножені.

Зверніть увагу, що ці ваги будуть помножені на `sample_weight` (переданий через метод `fit`), якщо вказано `sample_weight`.

ccp_alpha: *non-negative float, default=0.0* – Параметр складності, який використовується для скорочення мінімальної вартості та складності. `ccp_alpha` Буде вибрано піддерево з найбільшою вартістю складності, яка є меншою, ніж `ccp_alpha`. За замовчуванням обрізка не виконується.

Всі категоріальні змінні потрібно перевести в One-hot-encoded вигляд. Наприклад, замість колонки Lead мають бути колонки з назвами кожного лендінга та позначено «1», або «0», що означатиме чи належить цей користувач до цієї Landing Page, чи ні, відповідно.

Реалізація функції для One-hot-encoding категоріальних змінних, рисунок 2.16:

```
In [20]: #One-hot-encoding категоріальних змінних
def encode_categorical(df_train, df_test):
    test_start_ind = df_train.index[-1] #запам'ятали індекс, з якого починаються тестові дані
    concated_df = pd.concat([df_train, df_test], sort=False) #об'єднали навчальну вибірку з тестовою, для коректності
    #перетворення
    encoded_df = pd.get_dummies(concated_df, drop_first=True)#перетворили
    df_train_encoded, df_test_encoded = encoded_df.iloc[:test_start_ind+1, :], encoded_df.iloc[test_start_ind:, :] #роз'єднали
    return df_train_encoded, df_test_encoded
```

Рисунок 2.16 – Реалізація функції для One-hot-encoding

В результаті отримуємо даний результат, рисунок 2.17 :

In [22]: X_default

landing_page_CZ_axela	landing_page_CZ_axela	landing_page_CZ_axela	landing_page_DE_gemini_axela	City_Znojmo	OS_GNU/Linux	OS_OS_X	OS_Ubuntu	OS_Windows	OS_iOS	Device_type_Mobile_phone	Device_type_TV	Device_type_Tab
ha26@gmail.com	viktoriaSyhazka@gmail.com	аксела	FB мой	...	0	0	0	0	0	0	1	0
0	0	0	0	...	0	0	0	0	0	0	1	0
0	0	0	0	...	0	0	0	0	0	0	1	0
0	0	0	0	...	0	0	0	0	0	0	1	0
0	0	0	0	...	0	0	0	0	0	0	1	0
...
0	0	0	0	...	0	0	0	0	1	0	0	0

Рисунок 2.17 – Результат виконання функції

Для аналізу даних потрібно провести навчання системи, рисунок 2.18 :

```
In [28]: %%time
dt_clf.fit(X_default, y_default_full) #навчання

CPU times: total: 0 ns
Wall time: 21.4 ms

Out[28]: DecisionTreeClassifier
DecisionTreeClassifier()
```

Рисунок 2.18 – Аналіз даних

Для візуалізації аналізу потрібно імпортувати клас `export_graphviz` з бібліотеки `sklearn.tree` та класи `Image`, `display` з бібліотеки `IPython.display`. Також необхідно реалізувати функцію, яка буде приймати файл `.dot`, надавати його для обробки попереднім класам та відображати отриманий результат, рисунок 2.19 :

```
In [30]: from sklearn.tree import export_graphviz
from IPython.display import Image, display

def visualize_tree(clf, X, file_name):
    export_graphviz(clf, feature_names=X.columns, out_file=file_name + '.dot', filled=True)
    !dot -Tpng {file_name}.dot -o {file_name}.png
    display(Image(filename=file_name + '.png'))

In [31]: visualize_tree(dt_clf, X_default, 'tree');
```

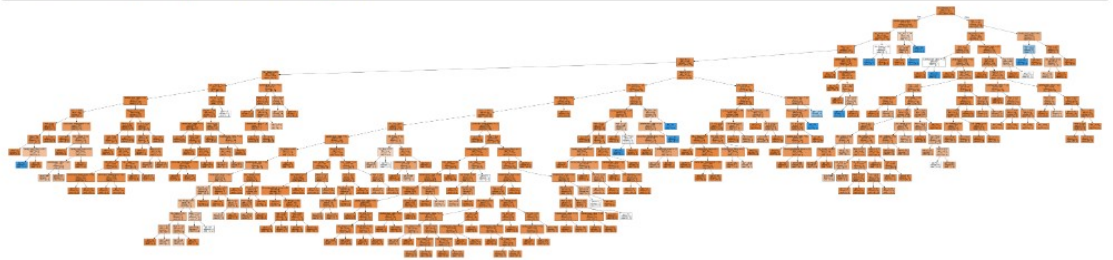


Рисунок 2.19 – Результат візуалізації аналізу

В результаті отримано повне, занадто велике дерево, яке є надто складним для аналізу. Для кращого розуміння та аналізу інформації вирішено зменшити глибину дерева та вказати деякі параметри для бажаного результату.

Із бібліотеки `sklearn.model_selection` імпортовано клас `GridSearchCV`.

GridSearchCV реалізує метод «fit» і «score». Він також реалізує «score_samples», «predict», «predict_proba», «decision_function», «transform» і «inverse_transform», якщо вони реалізовані у використовуваному оцінювачі.

Параметри засобу оцінки, який використовується для застосування цих методів, оптимізуються за допомогою перехресно перевіреного пошуку в сітці параметрів.

Параметри:

`estimator`: estimator object - передбачається, що це реалізує інтерфейс оцінювача scikit-learn. Або засіб оцінки повинен надати score функцію, або scoring його потрібно передати.

`param_grid` dict або список словників – словник із назвами параметрів (str) як ключі та списки налаштувань параметрів, які можна спробувати як значення, або список таких словників, у цьому випадку досліджуються сітки, охоплені кожним словником у списку. Це дозволяє здійснювати пошук по будь-якій послідовності налаштувань параметрів.

`scoring`: str, callable, list, tuple or dict, default=None

Стратегія оцінки ефективності перехресно перевіреної моделі на тестовому наборі.

Якщо `scoring` представляє одну оцінку, можна використовувати:

1. Один рядок.
2. Виклик, який повертає одне значення.

Якщо `scoring` представляє кілька оцінок, можна використовувати:

1. Список або кортеж унікальних рядків.
2. Викликається, що повертає словник, де ключами є імена метрик, а значеннями є оцінки метрик.
3. Словник з іменами метрик як ключами та значеннями, які можна викликати.

`n_jobs` int, default=None

Кількість завдань, які виконуються паралельно. None означає 1, якщо не в `joblib.parallel_backend` контексті. -1 означає використання всіх процесорів.

`refitbool, str, or callable, default=True`

Переналаштуйте оцінювач, використовуючи найкращі знайдені параметри для всього набору даних.

Для оцінки з кількома показниками це має бути `str` позначення оцінювача, який використовуватиметься для пошуку найкращих параметрів для переналаштування оцінювача в кінці.

Якщо під час вибору найкращого оцінювача враховуються інші міркування, окрім максимального балу, `refit` можна встановити функцію, яка повертає вибране `best_index_` задане `cv_results_`. У цьому випадку `best_estimator_` і `best_params_` буде встановлено відповідно до поверненого `best_index_`, тоді як `best_score_` атрибут буде недоступним.

Переобладнаний оцінювач стає доступним в `best_estimator_` атрибуті та дозволяє використовувати `predict` його безпосередньо в цьому `GridSearchCV` екземплярі.

Крім того, для оцінки кількох показників атрибуту `best_index_` та будуть доступні, лише якщо `best_score_` встановлено, і всі вони визначатимуться відповідно до цього конкретного показника `best_params_refit`.

Перегляньте `scoring` параметр, щоб дізнатися більше про оцінку кількох показників.

Щоб дізнатися, як розробити спеціальну стратегію вибору за допомогою функції виклику через `refit`.

`cvint, cross-validation generator or an iterable, default=None`

Визначає стратегію розподілу перехресної перевірки. Можливі вхідні дані для `cv`:

1. Немає, щоб використовувати стандартну 5-кратну перехресну перевірку.
2. Ціле число, щоб вказати кількість згорток у (Stratified)Kfold.
3. Ітерований результат (поїзд, тест) розбивається як масиви індексів.

Для цілочисельних/немає вхідних даних, якщо оцінювач є класифікатором є двійковим або мультикласовим StratifiedKFold. У всіх інших випадках KFold використовується. Ці розділювачі створено за допомогою `shuffle=False`, тому розділення буде однаковим для всіх викликів.

`Verboseint` – контролює багатослівність: чим вище, тим більше повідомлень.

1. `>1`: відображається час обчислення для кожного згортання та параметра-кандидата;
2. `>2`: також відображається оцінка;
3. `>3`: індекси параметрів згортання та кандидата також відображаються разом із часом початку обчислення.

`pre_dispatchint, or str, default='2*n_jobs'`

Контролює кількість завдань, які надсилаються під час паралельного виконання. Зменшення цього числа може бути корисним, щоб уникнути різкого споживання пам'яті, коли надсилається більше завдань, ніж ЦП може обробити. Цей параметр може бути:

1. Жодного, у цьому випадку всі завдання створюються та породжуються негайно. Використовуйте це для легких і швидких завдань, щоб уникнути затримок через створення завдань на вимогу
2. `Int`, що дає точну кількість створених вакансій
3. Рядок, який дає вираз як функцію `n_jobs`, як у «`2*n_jobs`»

`error_score'raise' or numeric, default=np.nan`

Значення, яке призначається оцінці, якщо під час підгонки оцінювача виникає помилка. Якщо встановлено значення 'raise', помилка виникає. Якщо задано числове значення, виникає `FitFailedWarning`. Цей параметр не впливає на крок оновлення, який завжди призведе до появи помилки.

`return_train_scorebool, default=False`

Якщо `False`, `cv_results_` атрибут не включатиме результати навчання. Обчислення навчальних балів використовується, щоб отримати уявлення про те, як різні налаштування параметрів впливають на компроміс із надмірним/недостатнім. Однак обчислення балів на навчальному наборі може бути обчислювально дорогим і не є суворо необхідним для вибору параметрів, які дають найкращу продуктивність узагальнення.

Потрібно вказати бажані параметри для класифікатора, рисунок 2.20 :

```
In [33]: clf = DecisionTreeClassifier()

parameters = {
    'criterion': ['entropy'],
    'max_depth': range(4, 10),
    'min_samples_split': range(5, 10),
    'min_samples_leaf': range(5, 10)
}

grid_cv = GridSearchCV(clf, parameters, cv=5)
```

Рисунок 2.20 – Критерій оцінювання було обрано за ентропією.

Ентропія — це не що інше, як міра безладу, або чистоти.

Математична формула ентропії така (2.41) :

$$E(S) = \sum_{i=1}^c -p_i \log_2 p_i \quad (2.41)$$

де p_i — це просто частотна ймовірність елемента/класу « i » в наших даних. Для простоти припустімо, що у нас є лише два класи, позитивний клас і негативний клас. Тому « i » тут може бути або $+$, або $-$. Отже, якби було загалом 100 точок даних у наборі даних, з яких 30 належали до позитивного класу, а 70 належали б

до негативного класу, тоді «P+» дорівнюватиме 3/10, а «P-» — 7/10. Досить просто.

Ентропія є мірою неупорядкованості або невизначеності, і метою моделей машинного навчання та Data Scientists загалом є зменшення невизначеності.

Максимальна глибина дерева обрана в діапазоні від 4 до 10. Мінімальна кількість зразків, необхідних для розбиття внутрішнього вузла обрано від 5 до 10. Мінімальна кількість зразків, необхідна для листкового вузла також обрана від 5 до 10.

Під час опрацювання даних процесор для обчислень використовувався протягом 10.8 секунд, а той час як весь процес зайняв 21.1 секунду, рисунок 2.21:

```
In [35]: best_model = grid_cv.best_estimator_  
  
In [36]: print('Кращі параметри класифікатора:', grid_cv.best_params_)  
Кращі параметри класифікатора: {'criterion': 'entropy', 'max_depth': 4, 'min_samples_leaf': 8, 'min_samples_split': 5}
```

Рисунок 2.21 – Параметри класифікатора

Система автоматично визначила, що серед запропонованих параметрів найкращим буде набір: {'criterion': 'entropy', 'max_depth': 4, 'min_samples_leaf': 8, 'min_samples_split': 5}, рисунок 2.22:

Отримано результат:

```
In [37]: grid_pred = best_model.predict(X_test)
```

```
In [38]: visualize_tree(best_model, X_default, 'tree_gridsearch')
```

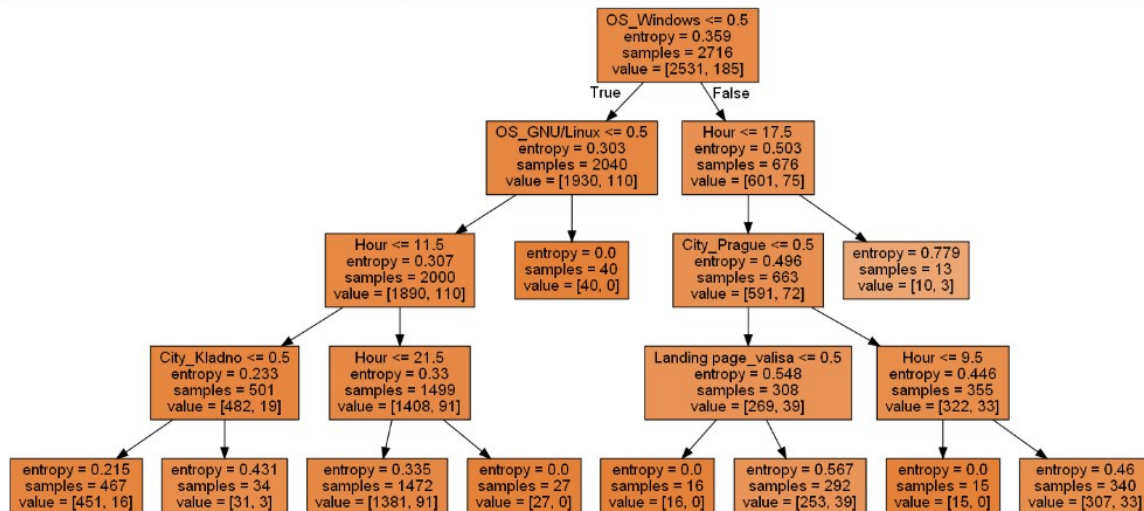


Рисунок 2.22 – Результат обробки даних

Для кращого бачення ситуації потрібно дати можливість системі в автоматичному режимі підібрати параметри та знайти результат. Для цього необхідно підключити клас `RandomizedSearchCV` бібліотеки `sklearn.model_selection`, рисунок 2.23:

```
In [46]: from sklearn.model_selection import RandomizedSearchCV
```

```
In [47]: random_grid_cv = RandomizedSearchCV(clf, parameters, cv=7)
```

```
In [48]: %%time
random_grid_cv.fit(X_default, y_default_full);
```

```
CPU times: total: 828 ms
Wall time: 1.51 s
```

```
Out[48]: RandomizedSearchCV
> estimator: DecisionTreeClassifier
> DecisionTreeClassifier
```

```
In [49]: #обираємо кращу модель з навченим
best_model_rand = random_grid_cv.best_estimator_
```

Рисунок 2.23 – Клас `RandomizedSearchCV` бібліотеки `sklearn.model_selection`

```
In [50]: print('Кращі параметри класифікатора:', random_grid_cv.best_params_)
```

```
Кращі параметри класифікатора: {'min_samples_split': 5, 'min_samples_leaf': 8, 'max_depth': 4, 'criterion': 'entropy'}
```

```
In [51]: rand_grid_pred = best_model_rand.predict(X_test)
```

```
In [52]: visualize_tree(best_model_rand, X_default, 'tree_rand')
```

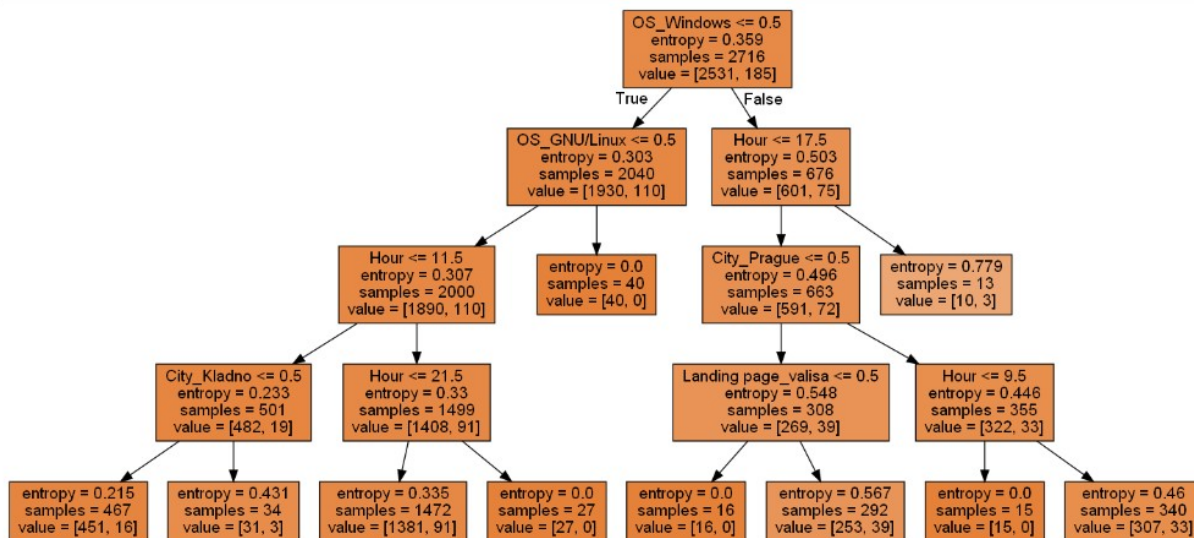


Рисунок 2.24 – Дерево класифікацій

Час відведений на випадкове створення дерева класифікацій значно менший за ручний підбір параметрів, але також має свої недоліки. Іноді інформація може бути не повною, чи погано структурованою для аналізу. Отримано результат у вигляді дерева, рисунок 2.24.

3 ПІДВИЩЕННЯ РЕЗУЛЬТАТИВНОСТІ РЕКЛАМНОЇ КОМПАНІЇ

3.1 Розробка методик підвищення результативності рекламної компанії

Для підвищення результативності рекламної компанії в магістерській роботі запропонована методика на основі методів ММ. А саме для кластеризації цільової аудиторії та її звичок. Запропоновано методика дозволяє зменшити фінансові витрати на рекламну компанію, чітко визначити цільову аудиторію, підвищити лояльність користувачів.

Методика складається з наступних етапів, рисунок 3.1 : 8/527

1. Підготовка до рекламної компанії, аналіз ринку.
2. Розробка web-сторінок для рекламної компанії.
3. Створення рекламних креативів.
4. Виконувати доки поки активна рекламна компанія:
 - 4.1 Для кожної з web-сторінок та креативів для них запуск реклами.
 - 4.2 Збір даних.
 - 4.3 Аналіз даних.
 - 4.4 Вибір цільової аудиторії.
 - 4.5 Вибір web-сторінок, креативів.
 - 4.6 Аналіз та доопрацювання web-сторінок, креативів.
5. Завершення рекламної компанії.

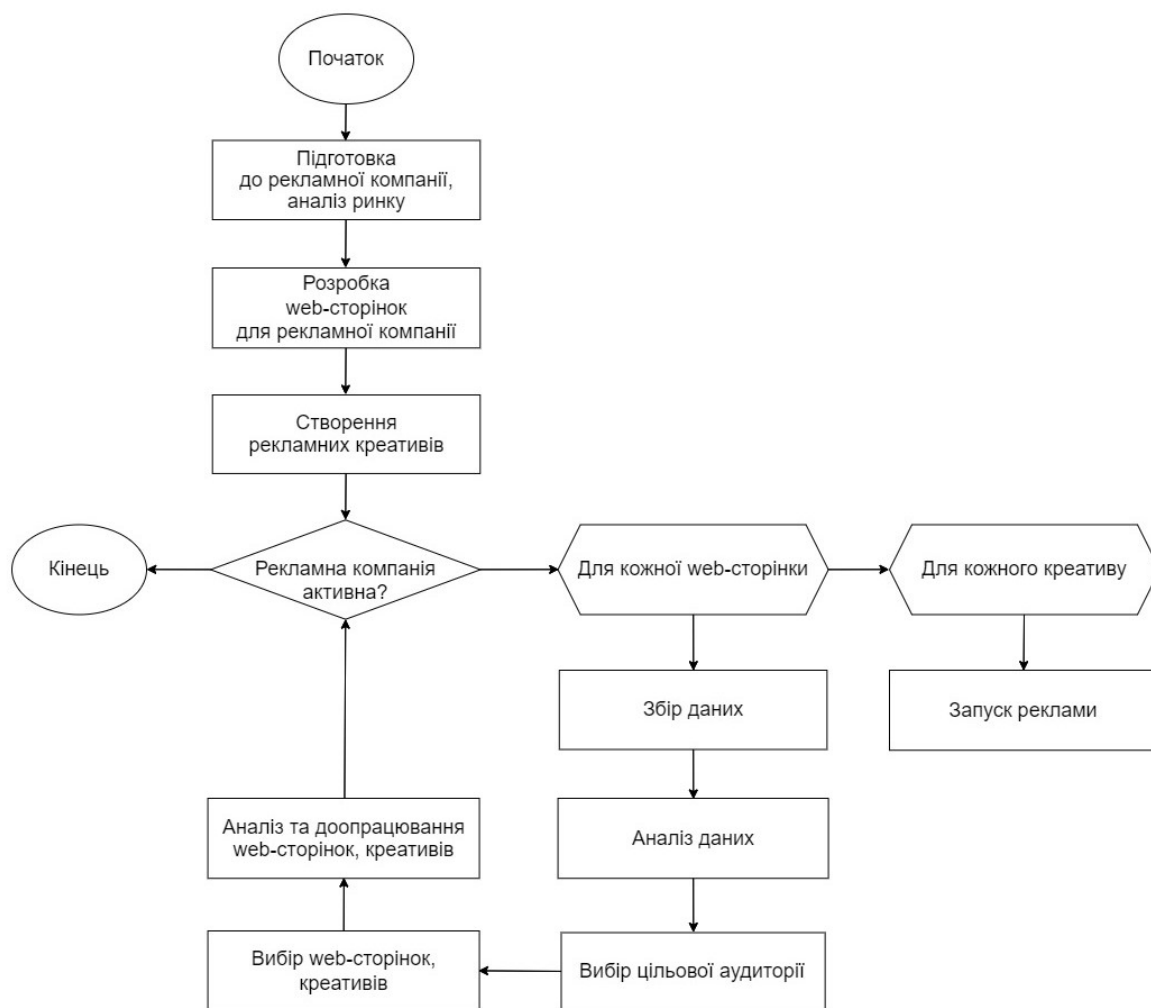


Рисунок 3.1 – Блок-схема результативності рекламної компанії

3.2 Аналіз ефективності запропонованої методики

Під час аналізу даних, отриманих в результаті роботи машинного навчання було помічено, що відсоток користувачів комп'ютерів, які відвідали сторінку та стали клієнтами рекламодавця набагато більший, аніж той самий показник у користувачів мобільних пристроїв. Варто зауважити, що це були користувачі саме Windows, які живуть в Празі, користувалися Інтернетом до 18:00, яким найбільше припала до увагу web-сторінка та креатив Landing page_valisa. Також 23% користувачів, які відвідали сторінку після 18:00 також стали клієнтами. Під ніч реклама не була для користувачів актуальною та відсоток лідів за цей період часу дорівнює 0. Після подальшого обговорення та перевірки було виявлено, що

розмір зображень на сайті рекламодавця занадто великий, а також сайт не повністю адаптовано для мобільних телефонів.

Було вирішено покращити User Interface та User Experience на сайті рекламодавця. Наступним кроком при запуску нової реклами був використаний креатив та web-сторінка Landing page_valisa в первинному стані та з деякими покращеннями. Час на показ реклами було обрано в межах з 09:00 до 12:00, а також з 15:00 до 22:00. При цьому хоч трафік в цій країні ввечері зазвичай дорожчий, аніж вдень, було вирішено сфокусуватися на користувачах, які користуються Інтернетом ввечері, оскільки вони мають вищий рівень лояльності.

Після проведеної роботи середній рівень лояльності користувачів підвищився з 5.77% до 20.70% при тих самих витратах на рекламу. А дохід рекламного агентства зріс на 312%.

Метод, розроблений в даній роботі, виконав поставлені задачі та успішно покращив рекламну компанію, збільшив дохід компанії та її авторитет, рисунок 3.2:

ID	Назва	Група	Джерело	Потоки	Кліки	Ліди	CR (ліди)	Продажі	CR (прод)	Конв.	CR	Дохід (підт)	Витрата	Прибуток	ROI
314	Test Machine Learning	Без групи		1	127 590	7 365	5,77%	2 702	2,12%	10 067	7,89%	\$107 925...	\$0	\$107 92...	0%
323	Second Test Machine Learning	Без групи		1	111 321	23 044	20,70%	11 238	10,09%	34 287	30,80%	\$444 879...	\$0	\$444 87...	0%
368	ДК оффер	Без групи		1	18 311	1 469	8,02%	268	1,46%	1737	9,49%	\$9 341,00...	\$0	\$9 341,00	0%
311	zalivun ninan direct link	Без групи		1	12 481	468	3,75%	209	1,67%	677	5,42%	\$8 340,00...	\$0	\$8 340,00	0%
409	валиун прилу	Без групи	LolaApps	1	336	95	28,27%	19	5,65%	114	33,93%	\$760,0000	\$0	\$760,00	0%
432	pmi offer br	Без групи		1	474	80	16,88%	15	3,16%	95	20,04%	\$450,0000	\$0	\$450,00	0%
375	ppc pinup	Без групи		1	1 341	50	3,73%	11	0,82%	61	4,55%	\$440,0000	\$0	\$440,00	0%
366	dk cl.pe offer link	Без групи		1	321	26	8,10%	10	3,12%	36	11,21%	\$400,0000	\$0	\$400,00	0%
415	олексій прила	Без групи	LolaApps	1	229	34	14,85%	0	0%	34	14,85%	\$0	\$0	\$0	0%
309	twinn app	Без групи		1	150	32	21,33%	0	0%	32	21,33%	\$0	\$0	\$0	0%
344	KZ Pinup	Без групи		1	741	14	1,89%	8	1,08%	22	2,97%	\$320,0000	\$0	\$320,00	0%
430	PM OFFER ZALIVUN	Без групи		1	58	8	13,79%	0	0%	8	13,79%	\$0	\$0	\$0	0%
429	apps cl zalivun	Без групи	LolaApps	1	19	2	10,53%	0	0%	2	10,53%	\$0	\$0	\$0	0%
414	pinup shk company	Shk		1	28	0	0%	1	3,57%	1	3,57%	\$10,0000	\$0	\$1,00	0%
416	organic br прила	Без групи	LolaApps	1	8	1	12,50%	0	0%	1	12,50%	\$0	\$0	\$0	0%
1	tetris1	Tetris	Facebook.com	1	4 492	0	0%	0	0%	0	0%	\$0	\$0	\$0	0%
2	Az Oper Most	Oper		1	1 359	0	0%	0	0%	0	0%	\$0	\$0	\$0	0%
4	2Garin Mostbet IN	2Garin	Facebook.com	1	0	0	0%	0	0%	0	0%	\$0	\$0	\$0	0%
5	2Garin Mostbet IN	2Garin	Facebook.com	1	0	0	0%	0	0%	0	0%	\$0	\$0	\$0	0%
				409	2 144 133	2 279	0,11%	541	0,03%	2 820	0,13%	\$20 052,0000	\$0	\$20 052,00	0%

Рисунок 3.2 – Вигляд рекламної компанії після покращення

3.3 Рекомендації для подальшого удосконалення рекламної компанії

Для успішної рекламної компанії потрібно не лише вміти запустити рекламу в Google Ads, цей процес вимагає висококваліфікованих фахівців, які працюють в команді, де кожен відповідає за свій процес та гарно в ньому знається. Проте інформаційні технології допомагають рекламі на кожному кроці. Починаючи від нотатки з новою ідеєю, закінчуючи звітом в електронній таблиці та балансом на екрані смартфона.

Сьогодні в рекламі активно використовують машинне навчання. Для майбутнього вдосконалення рекламних компаній було вирішено більш детально вивчити та застосовувати такі засоби, котрі створюють і редагують за вказаним описом фото та відео, а також покращують існуючі. На рисунку 3.3 відображено результат роботи нейронної мережі за запитом «Майбутня реклама в Україні».



Рисунок 3.3 – Результат роботи нейронної мережі

Наступним кроком до вдосконалення рекламної компанії є комбінація даного методу з іншими засобами, котрі пришвидшують та покращують роботу фахівців та робота над візуальною та функціональною частиною алгоритму, котрий було створено під час виконання даної магістерської роботи.

ВИСНОВКИ

В магістерській роботі запропоновано методику підвищення результативності рекламної компанії, що дозволяє підвищити рівень лояльності користувачів Інтернет та дохід рекламного агентства.

Мета магістерської роботи, як полягає в оптимізації пошуку цільової аудиторії для рекламної компанії на основі методів машинного навчання успішно досягнута.

В роботі проаналізовані методи машинного навчання такі, як: штучна нейронна мережа, дерево лінійної регресії та дерево регресії. Встановлено, що найменшу похибку для прогнозування характеристик цільової аудиторії дає метод дерева регресії. Тому саме він був використаний в дослідженні та успішно застосований на практиці.

Детально розглянутий таргетинг, його характеристики, переваги та недоліки. Все це лягало в основу методики вибору цільової аудиторії та дозволило побудувати методику підвищення результативності рекламної компанії.

Результати дослідження були використані для підвищення ефективності рекламних компаній та рівня лояльності користувачів Інтернет.

Наступним кроком до вдосконалення рекламної компанії є комбінація даного методу з іншими засобами, котрі пришвидшують та покращують роботу фахівців та робота над візуальною та функціональною частиною алгоритму, котрий було створено під час виконання даної магістерської роботи.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Данило А. І. Методика калькулювання собівартості на основі систем таргет-костингу та кайзен-костингу // Сучасний стан та перспективи розвитку обліку, аналізу, аудиту, звітності і оподаткування в умовах євроінтеграції: збірник наукових статей випускників з фаху «Облік і оподаткування» // Випуск 11 (2021). – Ужгород, 2021.- С. 35 - 42.
2. Павлова, С. І. Метод таргет-костингу як прогресивний метод цільового стратегічного управління витратами [Текст] / С. І. Павлова, Т. М. Магнушевська // Науковий вісник Ужгородського національного університету : Серія: Міжнародні економічні відносини та світове господарство / голов. ред. М.М. Палінчак. – Ужгород : Видавничий дім «Гельветика», 2016. – Вип. 10 Ч. 2. – С. 56–58. – Бібліогр.: с. 58 (10 назв).
3. Алейнікова О.В. Особливості використання режимів таргетування в сучасних умовах розвитку грошово-кредитної політики. Інвестиції: практика та досвід. 2013. № 14. С. 22–25.
4. Кулінець А. П. Особливості міжнародного досвіду використання режимів таргетування валютного курсу. Наукові праці НДФІ. 2008. Вип. 3. С. 160–168.
5. Семенова Ю.С. Таргетування інфляції: світовий досвід та можливості застосування в Україні. Формування ринкових відносин в Україні. 2008. № 1. С. 10–14.
6. Примостка Л. О. Перспективи впровадження інфляційного таргетування в Україні. Наукові записки Національного університету «Острозька академія». Серія «Економіка» : науковий журнал. Острог : Вид-во НУ«ОА», вересень 2017. № 6(34). С. 73–78.
7. Monden Y. and K. Hamada (1991) Target Costing and Kaizen Costing in Japanese Automobile Companies // Journal of Management Accounting Research, Vol. 3. — P. 16—34.

8. Кузьменко К.О. Коцепція таргет-костинг як дієвий метод стратегічного обліку // Збірник наукових праць науково-практичної конференції магістрів та студентів "Облік, аналіз і аудит: теорія, практика, перспективи". — Мелітополь, 2009. — № 9 — С. 179—183.
9. Романова, А.В. Таргетована реклама як ефективний спосіб просування в соціальних мережах [Текст] / А. В. Романова, З. М. Андрушкевич, О. Б. Вальков // Вісник Хмельницького національного університету. Економічні науки. — 2019. — №5. — С. 207-210.
10. Гриценко А.А. Інститут таргетування інфляції: зарубіжний досвід і перспективи запровадження в Україні: наукова доповідь / А.А. Гриценко, Т.О. Кричевська, О.І. Петрик. — Київ: Інститут економіки та прогнозування НАН України, 2008. — 271 с.
11. Дзюблюк О. Проблеми інфляційного таргетування в сучасній економіці / О. Дзюблюк // Світ фінансів. — 2015. — Вип. 2. — С. 7-18.
12. Ковальова Т.В. Таргетування інфляції: зарубіжний досвід та необхідність впровадження в Україні / Т.В. Ковальова, В.М. Бровко // Проблеми і перспективи розвитку підприємництва. — 2011. — № 1. — С. 125-128.
13. Міщенко В.І. Міжнародний досвід інфляційного таргетування: модель успіху для України [Електронний ресурс] / В.В. Іванов, В.І. Міщенко, О.К. Малютін // Актуальні проблеми економіки. — 2015. — № 4. — С. 414-425.
14. Семенова Ю.С. Таргетування інфляції: світовий досвід та можливості застосування в Україні / Ю.С. Семенова // Формування ринкових відносин в Україні. — 2008. — № 1. — С. 10-14.
15. Уткіна О. Необхідність таргетування інфляції в Україні / О. Уткіна // Економіст. — 2014. — № 12. — С. 19-22.
16. Данько, Н. (1). Брендінг як напрям вдосконалення рекламної стратегії туристичного підприємства. Вісник Харківського національного університету імені В. Н. Каразіна. Серія: Міжнародні відносини. Економіка. Країнознавство. Туризм, (1086), 133-136.

17. Яненко, Я.В. Інновації як основа для успішної реклами товарів і послуг [Текст] / Я.В. Яненко // Український інформаційний простір : науковий журнал Інституту журналістики і міжнародних відносин Київського національного університету культури і мистецтв / гол. редактор М.С. Тимошик. – Число 2. – К. : КНУКіМ, 2014. – С. 325–328.
18. Головчук, Ю. О., & Дибчук, Л. В. (2020). Формування та особливості ефективної рекламної кампанії. Економіка, управління та адміністрування, (1(91), 21–26.
19. Яненко, Я. В. Сучасні бренд-комунікації: соціалізаційний аспект / Я. В. Яненко // Образ : науковий збірник. – Київ : Інститут журналістики КНУ ім. Т. Шевченка, 2016. – Вип. 4 (22). – С. 97–102.
20. Зінюк, А. В., & Каторов, Д. О. (2021). Комунікативні індикатори контекстної реклами у системі цифрового маркетингу. «СОЦІОПРОСТІР: МІЖДИСЦИПЛІНАРНИЙ ЕЛЕКТРОННИЙ ЗБІРНИК НАУКОВИХ ПРАЦЬ З СОЦІОЛОГІЇ ТА СОЦІАЛЬНОЇ РОБОТИ», (11), 46-52.
21. Гольцман І.О. Особливості застосування реклами в туристичній індустрії / І.О Гольцман // Актуальні проблеми, сучасний стан та перспективи розвитку індустрії туризму в Україні та Польщі: Матеріали другої міжнародної науковопрактичної конференції / За ред. Саух І.В. – Житомир: Вид-во ЖФ КІБІТ, 2009 – С. 132-136.
22. Цуканова, В. Я. Удосконалення процесу планування маркетингових досліджень на підприємстві / В. Я. Цуканова, О. В. Логвіненко // Проблеми і перспективи розвитку підприємництва : зб. наук. пр. [Електронний ресурс] / Харків. нац. автомоб.-дор. ун-т. – Харків, 2013. – № 2 (5). – С. 69–73.
23. Головчук, Ю. О., & Дибчук, Л. В. (2020). Формування та особливості ефективної рекламної кампанії. Економіка, управління та адміністрування, (1(91), 21–26. [https://doi.org/10.26642/ema-2020-1\(91\)-21-26](https://doi.org/10.26642/ema-2020-1(91)-21-26).
24. Mitchell TM. Machine learning. New York: McGraw-Hill; 1997.
25. Alpaydin E. Introduction to machine learning. 3rd ed. Cambridge, MA: The MIT Press; 2014.

26. Bishop CM. Pattern recognition and machine learning. New York: Springer; 2006.
27. Apolloni B. Machine learning and robot perception. Berlin: Springer; 2005.
28. Ao S-I, Rieger BB, Amouzegar MA. Machine learning and systems engineering. Dordrecht/New York: Springer; 2010.
29. Gong Y, Xu W. Machine learning for multimedia content analysis. New York/London: Springer; 2007.
30. Yu J, Tao D. Modern machine learning techniques and their applications in cartoon animation research. 1st ed. Hoboken: Wiley; 2013.
31. Werbos PJ. Beyond regression: new tools for prediction and analysis in the behavioral sciences; PhD thesis, Harvard University, 1974.
32. Quinlan JR. Induction of decision trees. Mach Learn. 1986;1:81–106.
33. Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995;20:273–97.
34. Schapire RE. A brief introduction to boosting. In: Proceedings of the 16th international joint conference on artificial intelligence, vol. 2. Stockholm: Morgan Kaufmann Publishers Inc; 1999. p. 1401–6.
35. Breiman L. Random forests. Mach Learn. 2001;45:5–32.
36. Kargupta H. Next generation of data mining. Boca Raton: CRC Press; 2009.
37. W. Richert, L. P. Coelho, “Building Machine Learning Systems with Python”, Packt Publishing Ltd., ISBN 978-1-78216-140-0.
38. J. M. Keller, M. R. Gray, J. A. Givens Jr., “A Fuzzy KNearest Neighbor Algorithm”, IEEE Transactions on Systems, Man and Cybernetics, Vol. SMC-15, No. 4, August 1985.
39. P. Harrington, “Machine Learning in action”, Manning Publications Co., Shelter Island, New York, 2012.
40. Wright, R. E. (1995). Logistic regression. In L. G. Grimm & P. R. Yarnold (Eds.), Reading and understanding multivariate statistics (pp. 217–244). American Psychological Association.
41. Clemma J Muller, Richard F MacLehose, Estimating predicted probabilities from logistic regression: different methods correspond to different target populations,

International Journal of Epidemiology, Volume 43, Issue 3, June 2014, Pages 962–970, <https://doi.org/10.1093/ije/dyu029>.

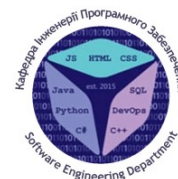
42. Ryan R. Pitre, Vesselin P. Jilkov, and X. Rong Li "A comparative study of multiple-model algorithms for maneuvering target tracking", Proc. SPIE 5809, Signal Processing, Sensor Fusion, and Target Recognition XIV, (25 May 2005); <https://doi.org/10.1117/12.609681>.
43. Nuno Amado, Joao Gama, and Fernando Silva. Parallel implementation of decision tree learning algorithms. In The 10th Portuguese Conference on Artificial Intelligence on Progress in Artificial Intelligence, Knowledge Extraction, Multi-agent Systems, Logic Programming and Constraint Solving, pages 6–13, December 2001.
44. Catherine L. Blake, Eamonn J. Keogh, and Christopher J. Merz. UCI repository of machine learning databases. University of California, Irvine, Dept. of Information and Computer Sciences, 1998.
45. Leon Bottou and Olivier Bousquet. The tradeoffs of large scale learning. In Advances in Neural Information Processing Systems, volume 20. MIT Press, Cambridge, MA, 2008.
46. Lindsay, M. Target discovery. Nat Rev Drug Discov 2, 831–838 (2003). <https://doi.org/10.1038/nrd1202>.
47. Chen R, Liu X, Jin S, Lin J, Liu J. Machine Learning for Drug-Target Interaction Prediction. Molecules. 2018; 23(9):2208. <https://doi.org/10.3390/molecules23092208>
48. J. Bergstra, O. Breuleux, F. Bastien, P. Lamblin, R. Pascanu, G. Desjardins, J. Turian, D. Warde-Farley, and Y. Bengio. Theano: a CPU and GPU math expression compiler. In Proceedings of the Python for Scientific Computing Conference (SciPy), June 2010.
49. Vences, N.A.; Díaz-Campo, J.; Rosales, D.F.G. Neuromarketing as an Emotional Connection Tool Between Organizations and Audiences in Social Networks. A Theoretical Review. Front. Psychol. 2020, 11, 1787.

- 50.Sano, Y.; Takayasu, H.; Havlin, S.; Takayasu, M. Identifying long-term periodic cycles and memories of collective emotion in online social media. *PLoS ONE* 2019, 14, e0213843.
- 51.Lee, N.; Broderick, A.J.; Chamberlain, L. What is “neuromarketing”? A discussion and agenda for future research. *Int. J. Psychophysiol.* 2006, 63, 199–204.
- 52.Hsu, L.; Chen, Y.-J. Neuromarketing, subliminal advertising, and hotel selection: An EEG study. *Australas. Mark. J.* 2020, 28, 200–208.
- 53.Hammou, K.A.; Galib, M.H.; Melloul, J. The Contributions of Neuromarketing in Marketing Research. *J. Manag. Res.* 2013, 5, 20–33.
- 54.Murphy, E.R.; Illes, J.; Reiner, P.B. Neuroethics of neuromarketing. *J Consum Behav.* 2008, 7, 293–302.
- 55.Renvoise, C.; Morin, P. *Neuromarketing: Understanding the "Buy Button" in Your Customer's Brain*; Thomas Nelson Publishers: Nashville, TN, USA, 2007.
- 56.M. W. Roth, "Survey of neural network technology for automatic target recognition," in *IEEE Transactions on Neural Networks*, vol. 1, no. 1, pp. 28-43, March 1990, doi: 10.1109/72.80203.
- 57.DARPA Neural Network Study, VA, Fairfax:AFCEA International Press, 1988.
- 58.B. Bhanu, "Automatic target recognition: State of the art survey", *IEEE Trans. Aerospace Electron. Syst.*, vol. AES-22, no. 4, pp. 364-379, July 1986.
- 59.M. Gyer, Automatic target recognition panel report, Dec. 1988.
- 60.S. Grossberg, "Nonlinear neural networks: Principles mechanisms and architectures", *Neural Networks*, vol. I, pp. 16-61, 1988.

ДОДАТОК А



ДЕРЖАВНИЙ УНІВЕРСИТЕТ ТЕЛЕКОМУНІКАЦІЙ
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ІНФОРМАЦІЙНИХ
ТЕХНОЛОГІЙ



Кафедра інженерії програмного забезпечення

МАГІСТЕРСЬКА РОБОТА

«ПІДВИЩЕННЯ РЕЗУЛЬТАТИВНОСТІ РЕКЛАМНОЇ КОМПАНІЇ НА ОСНОВІ МЕТОДІВ МАШИННОГО НАВЧАННЯ»

Виконав: Студент групи ПДМ-61 Яновський Д.А.

Керівник: дтн, доцент, завідувач кафедри Технологій цифрового зв'язку
Жебка В.В.

Київ - 2022

АКТУАЛЬНІСТЬ

1. Збільшення рівня цільової аудиторії та її лояльності спонукає збільшенню доходу та репутації компанії.
2. Машинне навчання надає можливість швидко обробляти великий об'єм даних, вміє самонавчатися та це допомагає фахівцям якісніше виконувати свою роботу.

МЕТА, ОБ'ЄКТ, ПРЕДМЕТ ДОСЛІДЖЕННЯ

Мета роботи: оптимізація пошуку цільової аудиторії для рекламної компанії на основі методів машинного навчання.

Об'єкт дослідження: процес кластеризації цільової аудиторії для таргету.

Предмет дослідження: методи машинного навчання.

ЗАВДАННЯ

1. Проаналізувати наукові праці з досліджуваної проблеми і обґрунтувати застосування машинного навчання для реклами.
2. З'ясувати особливості підвищення результативності рекламної компанії на основі методів машинного навчання.
3. Здійснити порівняльний аналіз методів кластерного аналізу: дерево рішень та бустерний аналіз.
4. Розробити та практично обґрунтувати методику підвищення результативності рекламної компанії на основі методів кластерного аналізу.
5. Розробити програмне забезпечення для кластеризації цільової аудиторії

МЕТОДИ ДОСЛІДЖЕННЯ

1. Методи машинного навчання.
2. Імітаційне моделювання.
3. Нейронна мережа.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Науковий результат: розроблено алгоритм кластеризації цільової аудиторії в рекламній сфері на основі методів машинного навчання.

Практичний результат: основні положення та результати магістерської роботи були використані для підвищення ефективності рекламних компаній та рівня лояльності користувачів Інтернет.

ДЕРЕВО РІШЕНЬ

Переваги	Недоліки
<ol style="list-style-type: none"> 1. Простота в розумінні та інтерпретації; 2. не вимагають підготовки даних; 3. дозволяють оцінити модель за допомогою статистичних тестів; дозволяють створювати класифікаційні моделі в тих областях, де аналітику досить складно формалізувати знання; 4. алгоритм конструювання дерева рішень не вимагає від користувача вибору вхідних атрибутів; швидко навчаються. 	<ol style="list-style-type: none"> 1. Складні конструкції; 2. недостатньо повно представляють дані.

Математичне формування дерева рішень:

$$Q_m^{left}(\theta) = \{(x, y) | x_j \leq t_m\}$$

$$Q_m^{right}(\theta) = Q_m \setminus Q_m^{left}(\theta)$$

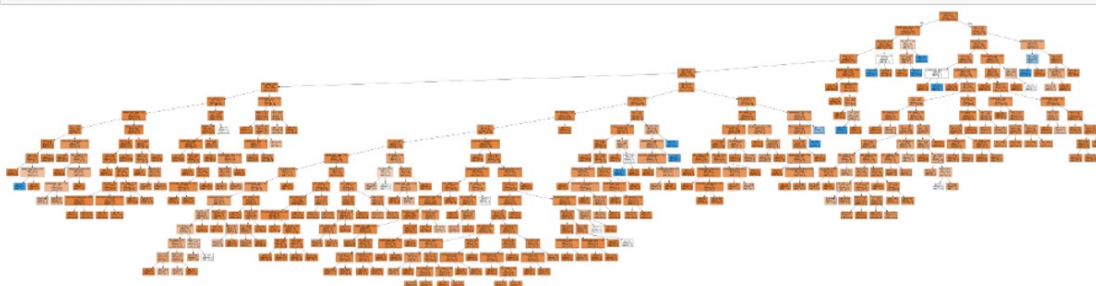
ДЕРЕВО РІШЕНЬ

1. Повне дерево

```
In [30]: from sklearn.tree import export_graphviz
from IPython.display import Image, display

def visualize_tree(clf, X, file_name):
    export_graphviz(clf, feature_names=X.columns, out_file=file_name + '.dot', filled=True)
    !dot -Tpng {file_name}.dot -o {file_name}.png
    display(Image(filename=file_name + '.png'))
```

```
In [31]: visualize_tree(dt_clf, X_default, 'tree');
```

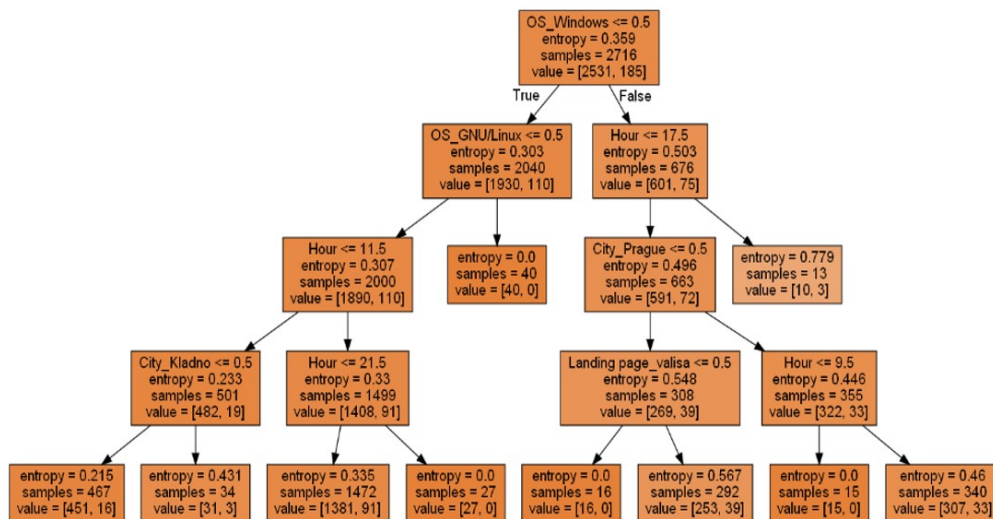


ДЕРЕВО РЕШЕНИЙ

2. Кращий результат

In [37]:
`grid_pred = best_model.predict(X_test)`

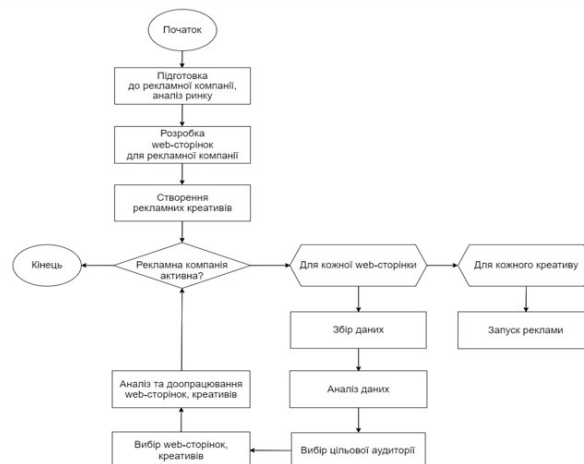
In [38]:
`visualize_tree(best_model, X_default, 'tree_gridsearch')`



МЕТОДИКА ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ РЕКЛАМНОЇ КОМПАНІЇ

Нейронні мережі — це набір алгоритмів, створених за зразком людського мозку, призначених для розпізнавання шаблонів. Вони інтерпретують сенсорні дані за допомогою машинного сприйняття, маркування або кластеризації вихідних даних.

Структура методики підвищення ефективності рекламної компанії



Алгоритми контрольованого машинного навчання

Логістична регресія
Дерево рішень
Штучні нейронні мережі

Алгоритми неконтрольованого машинного навчання

Факторний аналіз
Дискримінантний аналіз
Кластерний аналіз

ПЕРЕВАГИ ТА НЕДОЛІКИ ВИКОРИСТАННЯ НЕЙРОННИХ МЕРЕЖ В РЕКЛАМІ

Переваги	Недоліки
<ol style="list-style-type: none"> 1. Можливості паралельної обробки означають, що мережа може виконувати декілька завдань одночасно. 2. Інформація зберігається у всій мережі, а не лише в базі даних. 3. Здатність вивчати та моделювати нелінійні, складні зв'язки допомагає моделювати реальні зв'язки між входом і виходом. 4. Відмовостійкість означає, що пошкодження однієї чи кількох комірок ШНМ не призведе до зупинки генерації виводу. 5. Здатність створювати результати з неповними знаннями з втратою продуктивності залежно від того, наскільки важливою є відсутня інформація. 6. Жодних обмежень не кількість вхідних змінних. 7. Машинне навчання означає, що ШНМ може вчитися на подіях і приймати рішення на основі спостережень. 8. Здатність вивчати приховані зв'язки в даних без встановлення будь-яких фіксованих зв'язків означає, що ШНМ може краще моделювати дуже мінливі дані та непостійну дисперсію. 9. Здатність узагальнювати та виводити невидимі зв'язки з невидимими даними означає, що ШНМ можуть передбачати вихід невидимих даних. 	<ol style="list-style-type: none"> 1. Відсутність правил для визначення правильної структури мережі означає, що відповідну архітектуру штучної нейронної мережі можна знайти лише шляхом проб, помилок і досвіду. 2. Потреба в процесорах із можливостями паралельної обробки робить нейронні мережі апаратно залежними. 3. Мережа працює з числовою інформацією, тому всі проблеми повинні бути переведені в числові значення, перш ніж їх можна буде представити ШНМ. 4. Відсутність пояснень у рішеннях зондування є одним із найбільших недоліків ШНМ. Неможливість пояснити, чому або як стоїть за рішенням, породжує брак довіри до мережі.

РЕЗУЛЬТАТ ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ РЕКЛАМНОЇ КОМПАНІЇ

Keitaro BUSINESS Доступне оновлення!

Дашборд Компанії Лендінги Партнерські мережі Офери Джерела Звіти Тренди Домени

<input type="checkbox"/>	ID	Назва	Група	Джерело	Потоки	Кліки	Ліди	CR (ліди)	Продажі	CR (прод)	Конв.	CR	Дохід (підт)	Витрата	Прибуток	ROI	
<input checked="" type="checkbox"/>	314	Test Machine Learning	Без групи		1	127 590	7 365	5,77%	2 702	2,12%	10 067	7,89%	\$107 925,...	\$0	\$107 92...	0%	↻ 📄 ⋮
<input checked="" type="checkbox"/>	323	Second Test Machine Learning	Без групи		1	111 321	23 044	20,70%	11 238	10,09%	34 282	30,80%	\$444 879,...	\$0	\$444 87...	0%	↻ 📄 ⋮
<input type="checkbox"/>	368	ДК оффер	Без групи		1	18 311	1 469	8,02%	268	1,46%	1737	9,49%	\$9 341,00...	\$0	\$9 341,00	0%	↻ 📄 ⋮
<input type="checkbox"/>	311	zaliveu nianan direct link	Без групи		1	12 481	468	3,75%	209	1,67%	677	5,42%	\$8 340,00...	\$0	\$8 340,00	0%	↻ 📄 ⋮
<input type="checkbox"/>	409	zaliveu прилу	Без групи	LolaApps	1	336	95	28,27%	19	5,65%	114	33,93%	\$760,0000	\$0	\$760,00	0%	↻ 📄 ⋮
<input type="checkbox"/>	ID	Назва	Група	Джерело	Потоки	Кліки	Ліди	CR (ліди)	Продажі	CR (прод)	Конв.	CR	Дохід (підт)				
<input checked="" type="checkbox"/>	314	Test Machine Learning	Без групи		1	127 590	7 365	5,77%	2 702	2,12%	10 067	7,89%	\$107 925,...				
<input checked="" type="checkbox"/>	323	Second Test Machine Learning	Без групи		1	111 321	23 044	20,70%	11 238	10,09%	34 282	30,80%	\$444 879,...				
<input type="checkbox"/>	368	ДК оффер	Без групи		1	18 311	1 469	8,02%	268	1,46%	1737	9,49%	\$9 341,00...				
<input type="checkbox"/>	430	PM OFFER ZALIVUN	Без групи		1	58	8	13,79%	0	0%	8	13,79%	\$0	\$0	\$0	0%	↻ 📄 ⋮
<input type="checkbox"/>	429	apps cl zaliveu	Без групи	LolaApps	1	19	2	10,53%	0	0%	2	10,53%	\$0	\$0	\$0	0%	↻ 📄 ⋮
<input type="checkbox"/>	414	pinup shk company	Shk		1	28	0	0%	1	3,57%	1	3,57%	\$1,0000	\$0	\$1,00	0%	↻ 📄 ⋮
<input type="checkbox"/>	416	organic br прила	Без групи	LolaApps	1	8	1	12,50%	0	0%	1	12,50%	\$0	\$0	\$0	0%	↻ 📄 ⋮
<input type="checkbox"/>	1	tetris1	Tetris	Facebook.com	1	4 492	0	0%	0	0%	0	0%	\$0	\$0	\$0	0%	↻ 📄 ⋮
<input type="checkbox"/>	2	Az Oper Most	Oper		1	1 359	0	0%	0	0%	0	0%	\$0	\$0	\$0	0%	↻ 📄 ⋮
<input type="checkbox"/>	4	2Garin Mostbet IN	2Garin	Facebook.com	1	0	0	0%	0	0%	0	0%	\$0	\$0	\$0	0%	↻ 📄 ⋮
<input type="checkbox"/>	5	2Garin Mostbet IN	2Garin	Facebook.com	1	0	0	0%	0	0%	0	0%	\$0	\$0	\$0	0%	↻ 📄 ⋮
					409	2 144 133	2 279	0,11%	541	0,03%	2 820	0,13%	\$20 052,0000	\$0	\$20 052,00	0%	

ВИСНОВКИ

У ході проведеного дослідження були вирішені всі поставлені завдання і відповідно до мети отримані наступні **результати**:

1. Проаналізовано наукові праці з досліджуваної проблеми, обґрунтовано застосування машинного навчання для оптимізації рекламних компаній.
2. З'ясовано особливості параметрів для кластеризації цільової аудиторії на основі методів машинного навчання.
3. Досліджено метод дерева рішень як найбільш ефективний метод для машинного навчання для таргетингу.
4. Розроблено і практично випробувано методіку для підвищення результативності рекламної компанії на основі методу дерева рішень.
5. Розроблено програмне забезпечення для кластеризації цільової аудиторії

ПУБЛІКАЦІЇ ТА АПРОБАЦІЯ РОБОТИ

Статті:

1. Яновський Д.А. Методика підвищення результативності рекламної компанії на основі методів машинного навчання// ТІТ. №3, 2022, с. 25-30

Тези доповідей:

1. Яновський Д.А. Розробка методики вибору цільової аудиторії для підвищення результативності рекламної компанії // XV Науково-технічна конференція «Сучасні інфокомунікаційні технології» . – Київ: ДУТ, 2022. – с. 93-95

ДЯКУЮ ЗА УВАГУ!